

Characterisation of a methyl-CpG binding domain  
protein in *Caenorhabditis elegans* and mammals

Michael J. Carr

Thesis presented for the Degree of  
Doctor of Philosophy

The University of Edinburgh  
1999



## **Declaration**

**I declare that this thesis was composed by myself and that the work is my own except where otherwise indicated.**

## Acknowledgments

Many thanks to all the members of the Bird lab, past and present and especially to Adrian for giving me the opportunity and then supporting a project where I worked on a number of different systems. Also to Susan Tweedie and Donald Macleod for listening to problems however trivial from day one, Brian Hendrich for lots of help particularly with the localisation of the mPBM1 protein and for identifying the worm gene that formed the basis of my project, Martin Simmen for invaluable assistance with graphics and Christine Struthers and Aileen Grieg for a lot of laughs over the years.

I would especially like to thank Karen Thijssen and Ronald Plasterk in Amsterdam for welcoming me so warmly during my stay at the NKI and introducing me to worm biology. To Mark Blaxter and Peter Hunt and everyone in the Blaxter lab in Zoology, Edinburgh for their resources and advice and Shelia Boyle and Wendy Bickmore at the Western General Hospital, Edinburgh for helping me with mapping.

Thanks to Ashley Robertson for a few beers and that special memory of the Nou Camp. Thanks also to Ray, Pete, Pree and Emma for keeping me almost sane during the write up and Dion for teaching me a little Capoeira. To Alex, Susie and Tanya for all the fun from Wonkys to the Clachaig and an especially big thank you to Brian O' Neill in Dublin for advice and encouragement over the years.

Finally, to my family for helping me get through the last four years and never doubting this day would finally come.

## Abstract

Vertebrate DNA is globally modified by the incorporation of 5-methylcytosine ( $m^5C$ ) exclusively within CpG dinucleotides. In mammals DNA methylation is negatively correlated with the transcriptional activity of genes. A family of proteins have been identified that are targeted to methylated DNA by a methyl-CpG binding domain (MBD) and serve to mediate the transcriptional silencing and potentially other biological consequences of cytosine methylation. In contrast to vertebrates the genomes of the invertebrates *Drosophila melanogaster* and *Caenorhabditis elegans* lack detectable methylation. The recently sequenced genome of the model organism *C. elegans* was found to contain a gene that potentially encoded a protein containing a methyl-CpG binding domain and intriguingly, two motifs present in chromatin-associated transcriptional regulators - the PHD-zinc finger and bromodomain. The gene was given the standard name *pbm-1* for PHD-zinc finger, bromodomain and MBD gene-1. The MBD of *pbm-1* was found to be highly homologous to a human EST suggesting evolutionary and functional conservation of this domain. A mouse cDNA, mPBM1 was identified that encoded a protein with the same modular organisation of motifs as the *C. elegans* PBM-1 protein. The aim of this project was the characterisation of this gene family using a cross-species approach that exploited the advantages particular to the nematode and mammalian systems.

The MBDs of PBM-1 and mPBM1 did not associate with methylated sequences in a series of in vitro binding studies. These findings do not rule out the possibility that the MBDs mediate binding to other covalent modifications of DNA or specific sequence targets. In support of the in vitro findings, the mPBM1 protein did not associate with the heavily methylated mouse major satellite when fused to GFP and was found to localise in a diffuse, granular pattern in the nucleus. Furthermore, the MBD of mPBM1 was dispensable for the spatial localisation of the protein. Deletion of a C-terminal element encompassing the PHD-zinc finger and bromodomain regions abolished the wild type localisation in mouse nuclei.

The PHD-zinc finger has been identified in a number of chromosomal translocations involved in acute leukaemias suggesting that deregulation of normal expression contributes to the transformation phenotype. The hPBM1 gene was physically mapped to chromosome 2q23 in a region found deleted and rearranged in leukaemias. Interestingly, mPBM1 fused to a heterologous DNA-binding domain was found to modestly repress the transcription of a reporter gene in mouse cells.

The field of DNA methylation has been unable to utilise the expertise present in the nematode, *C. elegans* because this organism appears to lack  $m^5C$ . An EMS-induced deletion within the *pbm-1* gene was identified and a *C. elegans* strain, *pbm-1* (*qa1700*) was established that was homozygous mutant for this deletion. The *pbm-1* (*qa1700*) strain was viable and the phenotype was indistinguishable from wildtype.



## Abbreviations

ATP	adenosine triphosphate
5-azaC	5-aza-2'-deoxycytidine
bp	base pair
BSA	Bovine serum albumin
cDNA	complementary deoxyribonucleic acid
Ci	Curies
Da	Dalton
DAPI	4, 6-diamino-2-phenylindole
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DEPC	diethylpyrocarbonate
dGTP	deoxyguanosine triphosphate
dH <sub>2</sub> O	distilled water
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
dUTP	deoxyuridine triphosphate
EDTA	ethylenediaminetetra-acetic acid
EMS	ethyl methanesulphonate
EST	expressed sequence tag
FISH	fluorescence <i>in situ</i> hybridisation
FITC	fluorescein isothiocyanate
GFP	Green Fluorescent Protein
GST	glutathione-S-transferase
HEPES	N-(Hydroxyethyl)piperazine-N'-[2-ethanosulphonic acid]
hr	hour
IPTG	isopropylthio- $\beta$ -D-galactoside
kb	kilobase
MBD	methyl-CpG binding domain
m <sup>5</sup> C	5-methylcytosine
min	minute
MOPS	3-(N-morpholino)propane-sulphonic acid
mPBM1	murine PHD-zinc finger, bromodomain and MBD1 protein
mRNA	messenger ribonucleic acid
MTase	DNA (cytosine-5) methyltransferase
OD	optical density

ONPG	orthonitrophenyl- $\beta$ -D-galactopyranoside
ORF	open reading frame
PCR	polymerase chain reaction
<i>pbm-1</i>	PHD-zinc finger, bromodomain and MBD gene-1
PBS	phosphate buffered saline
PMSF	phenylmethylsulphonylfluoride
PNK	polynucleotide kinase
RNA	ribonucleic acid
RNase	ribonuclease
RPM	revolutions per minute
RT	reverse transcriptase
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SSC	sodium chloride/sodium citrate
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
T/E	Tris EDTA
TMP	trimethylpsoralen
UTR	untranslated region
UV	ultraviolet
V	volts
w/v	weight for volume
X-gal	5-bromo-4-chloro-3-indoyl- $\beta$ -D-galactopyranoside

# Table of Contents

Title	i
Declaration	ii
Acknowledgments	iii
Abstract	iv
Abbreviations	v
Table of Contents	vii
 <b>Chapter One - Introduction</b>	 <b>1</b>
1.1 Introduction	2
1.1.1 DNA methylation in early embryogenesis	3
1.1.2 Maintenance and de novo DNA (cytosine-5) methyltransferases	4
1.1.3 Intrinsic mutability of 5-methylcytosine	5
1.2 DNA methylation outside the vertebrate lineage	6
1.2.1 Plant DNA methylation	6
1.2.2 Fungal DNA methylation	7
1.2.3 Invertebrate DNA methylation	8
1.3 DNA methylation and development	10
1.3.1 Reduced genomic methylation and developmental abnormalities	11
1.3.2 Phenotype of the <i>Dnmt1</i> knockout mouse	12
1.4 Molecular mechanisms for methylation-mediated repression	14
1.4.1 Direct inhibition of transcription factor binding	14
1.4.2 Indirect inhibition of transcription factor binding	14
1.5 Histone acetylation and deacetylation in gene control	16
1.5.1 Introduction	16
1.5.2 Histone acetylation and gene activation	17
1.5.3 Histone acetyltransferases	18
1.5.4 Histone deacetylases	21
1.6 Methyl CpG-binding Proteins (MeCPs)	24
1.6.1 The methyl CpG-binding domain (MBD)	26
1.6.2 MeCP2 is a transcriptional repressor and is essential for mouse development	26
1.7 The MBD protein family	28
1.7.1 Comparative genomics and expressed sequence tags (ESTs)	28

1.7.2 Employing ESTs to identify novel MBD-containing proteins	30
1.8 DNA methylation and histone deacetylation	31
1.8.1 DNA methylation recruits histone deacetylation	32
1.8.2 Chromatin remodeling complexes and DNA methylation	34
1.8.3 Conservation of the NuRD complex during evolution	37
1.8.4 Reversal of methylation-mediated silencing	38
1.9 MeCPs and expression inheritance	40
<b>Chapter Two - Materials and Methods</b>	<b>42</b>
2.1 Bacterial Cell Culture	43
2.1.1 Media and solutions	43
2.1.2 Bacterial strains	44
2.1.3 Preparation of competent cells for electroporation	44
2.1.4 Electroporation of competent cells	45
2.2 Cell and Tissue Culture	45
2.2.1 Media and reagents	45
2.2.2 Mammalian cell lines	46
2.2.3 Thawing of cells stored in liquid nitrogen	46
2.2.4 Maintenance of cell lines	46
2.2.5 Transfection of mammalian cell lines	47
2.2.6 $\beta$ -galactosidase assay	48
2.3 Preparation of Nucleic Acids	49
2.3.1 Isolation of plasmid DNA	49
2.3.2 Isolation of total RNA from cultured cells and tissues	50
2.3.3 Synthesis of cDNA for RT-PCR	50
2.4 Gel Electrophoresis	51
2.4.1 Electrophoresis of DNA	51
2.4.2 Electrophoresis of RNA	51
2.5 Radioisotopic labeling of DNA fragments	52
2.5.1 Random Primed Labeling of DNA probes	52
2.5.2 End Labeling of DNA oligonucleotides	52
2.6 Nucleic Acid Blotting and Hybridisation	53
2.6.1 Southern blot analysis of DNA	53
2.6.2 Northern Blot Analysis of RNA	54
2.6.3 Bacteriophage plaque lifts	54
2.6.4 Removal of Probe Following Hybridisation	55

2.7	Enzymatic Manipulation of DNA	55
2.7.1	Restriction endonuclease digestion of DNA	55
2.7.2	Dephosphorylation of DNA	55
2.7.3	Ligation of DNA	55
2.8	Polymerase Chain Reaction	56
2.9	DNA Sequencing	56
2.10	Screening of cDNA libraries	57
2.10.1	Preparation of phage-competent cells and phage isolation	57
2.10.2	In vivo excision of cDNA inserts from $\lambda$ ZAP II cDNA libraries	58
2.11	Fluorescent <i>In Situ</i> Hybridisation (FISH)	59
2.11.1	Labeling genomic probes with biotin-16-dUTP	59
2.11.2	Detection of biotin-16-dUTP incorporation	59
2.11.3	<i>In situ</i> hybridisation with biotin-labeled probes	60
2.12	Expression and affinity purification of recombinant proteins	61
2.12.1	Affinity purification of histidine tagged recombinant proteins	61
2.12.2	Affinity purification of glutathione-S-transferase (GST)-tagged recombinant proteins	62
2.13	DNA binding assays	63
2.13.1	Electrophoretic mobility shift assay	63
2.13.2	Southwestern assay	63
2.14	<i>Caenorhabditis elegans</i> handling, culture and genetic methods	65
2.14.1	Handling of Nematodes	65
2.14.2	Isolation of <i>C. elegans</i> insertion and deletion mutants	65
2.14.3	Outcrossing of <i>C. elegans</i> deletion mutants	67
2.14.4	PCR genotyping of nematodes	68
<b>Chapter Three - Identification of a family of <u>PHD</u>-zinc finger, <u>bromodomain</u> and <u>methyl-CpG</u> binding (PBM)-containing proteins</b>		69
3.1	Introduction	70
3.2	Results	70
3.2.1	Identification of the MBD-containing gene, <i>pbm-1</i> in the nematode <i>Caenorhabditis elegans</i>	70
3.2.2	Characterisation of the genomic structure and cDNA encoding PBM-1	71
3.2.3	Isolation of mPBM1 - a mammalian ortholog of <i>C. elegans</i> PBM-1	83

3.2.4 Structure and splice variants of the mPBM1 cDNA	97
3.2.5 Identification of a family of mammalian PBM proteins	103
3.2.6 Construction of bacterial expression vectors and the purification of PBM-1, mPBM1 and MBD4 MBDs	114
3.2.7 Analysis of the methyl-CpG binding and DNA binding activity of the <i>C. elegans</i> PBM-1-MBD and mouse PBM1-MBDs	123
3.3 Conclusions	133
<b>Chapter Four - Localisation and transcriptional regulatory properties of the mPBM1 protein</b>	<b>135</b>
4.1 Introduction to mPBM1 localisation	136
4.2 Results	136
4.2.1 Nuclear localisation of a full length GFP-mPBM1 fusion	136
4.2.2 Punctate localisation of a GFP-mPBM1 $\Delta$ MBD	140
4.2.3 Perturbed localisation of a GFP-mPBM1 $\Delta$ PHD/BrD	144
4.3 Conclusions	147
4.4 Transcriptional regulatory properties of mPBM-1	149
4.5 Results	150
4.6 Conclusions	155
<b>Chapter Five - Physical mapping of the hPBM1 gene</b>	<b>156</b>
5.1 Introduction to mapping of human PBM1	157
5.1.1 Chromosomal translocations involving transcriptional regulators	157
5.1.2 Chromosomal translocations and acetylases/deacetylases	157
5.1.3 Chromosomal translocations at 11q23 - the ALL1 locus	160
5.1.4 Chromosomal rearrangements of ALL1 - the PHD-zinc finger	160
5.2 Results	162
5.2.1 Mapping of hPBM1 using monochromosomal cell hybrids	162
5.2.2 Mapping of hPBM1 using fluorescence <i>in situ</i> hybridisation (FISH)	164
5.2.3 Mapping of hPBM1 by radiation hybrid mapping	167
5.3 Conclusions	167
<b>Chapter Six - Isolation of a <i>pbm-1</i> null mutant</b>	<b>168</b>
6.1 Introduction	169
6.2 Results	175
6.2.1 Identification of Tc1 transposon insertions in <i>pbm-1</i>	175
6.2.2 Identification and isolation of a <i>pbm-1</i> deletion mutant	182
6.2.3 Outcrossing of the <i>C. elegans</i> deletion mutant	193
6.3 Conclusions	197

<b>Chapter Seven - Discussion</b>	198
7.1 Summary of Results	199
7.2 The PHD-zinc finger, Bromodomain and MBD (PBM) protein family	201
7.3 The MBD regions of the PBM proteins do not bind methylated DNA	203
7.4 Localisation of the mPBM1 protein	206
7.5 Mapping of the hPBM1 gene to 2q23	210
7.6 Transcriptional regulatory properties of the mPBM1 protein	212
7.7 Potential protein interaction targets of the PBM proteins	214
7.8 The <i>C. elegans pbm-1</i> gene is not essential for viability	217
<b>References</b>	220

# CHAPTER ONE: INTRODUCTION



## 1.1 Introduction

The base 5-methylcytosine ( $m^5C$ ) represents about 1% of the nucleotides in the vertebrate genome. The covalent modification of DNA by methylation represents an epigenetic mechanism to modulate the informational content present within the primary sequence of nucleotide bases. The predominant methylated sequence in all animals is the self complementary dinucleotide CpG and the vertebrate genome is characterised by a bimodal pattern of DNA methylation where most (60-90%) of the CpGs are methylated with the remaining CpG residues conspicuously unmethylated (Gruenbaum et al., 1981a; Cooper et al., 1983). This small fraction of unmethylated CpG dinucleotides are present in discrete clusters, termed CpG islands that are found primarily at the 5' end of genes (Bird, 1986). In somatic tissues, CpG islands are constitutively unmethylated with certain exceptions such as the island-containing genes on the inactive X chromosome and parentally imprinted genes (Riggs and Pfeifer, 1992; Ferguson-Smith et al., 1993). Direct evidence exists that in vitro methylation of DNA and transfection into mammalian cells is correlated with transcriptional suppression (Busslinger et al., 1983; Yisraeli et al., 1988). However, methylation *per se* does not inhibit transcription as in certain cases silencing precedes methylation (Lock et al., 1987). Furthermore, methylated reporters are not immediately silenced when introduced to cells (Buschhausen et al., 1987) and loss of transcriptional activity coincides with the formation of chromatin (Kass et al., 1997). As vertebrates are characterised by genome-wide methylation this has suggested that DNA methylation might be involved in long term transcriptional silencing.

### 1.1.1 DNA methylation in early embryogenesis

The methylation of DNA in gametes is comparable to that found in adult tissues, however studies of embryonic methylation in the mouse have suggested the process is dynamic with a demethylation event occurring during pre-implantation development (Monk et al., 1987; Howlett and Reik, 1991). A wave of de novo methylation at around the time of implantation is thought to establish the adult modification pattern comprising genome wide methylation and constitutively unmethylated CpG islands (Shemer et al., 1991; Kafri et al., 1992). CpG islands were found to be resistant to this de novo methylation activity in transgenic mice studies and in cultured embryonic cells (Kolsto et al., 1986; Szyf et al., 1990; Frank et al., 1991). A passive mechanism of demethylation has been proposed where prevention of methyltransferase binding to a transcriptionally active CpG island results in demethylation (Matsuo et al., 1998). This is a replication-based model where two rounds of replication of symmetrically methylated DNA can result in a passive demethylation in the absence of maintenance methylation. Evidence does exist suggesting CpG islands in embryonic cells are transcriptionally active (Macleod et al., 1998) and asymmetric Sp1 elements within the CpG island of the *aprt* gene are also required to maintain the immunity of this region from the embryonic de novo methylation event (Macleod et al., 1994; Brandeis et al., 1994).

### 1.1.2 Maintenance and de novo DNA (cytosine-5) methyltransferases

The incorporation of a methyl group at position-5 of the cytosine ring is carried out by a conserved family of DNA (cytosine-5) methyltransferases. DNA methylation patterns are maintained by the action of Dnmt1, that preferentially methylates hemi-methylated DNA (Gruenbaum et al., 1982; Bestor and Ingram, 1983). The Dnmt1 enzyme is termed the maintenance DNA (cytosine-5) methyltransferase although the enzyme can also methylate unmethylated DNA in vitro (de novo methylation) suggesting a single enzymatic activity might mediate both maintenance and de novo functions. However, no de novo methyltransferase activity has been reported for Dnmt1 in vivo and ES cells homozygous for a null mutation in the *Dnmt1* gene were able to de novo methylate a provirus introduced into these cells (Lei et al., 1996). This suggested that additional activities independent of the Dnmt1 enzyme exist in mouse and these enzymes represent strong candidates for mediating the de novo methylation of genomic DNA occurring at the time of implantation during mouse development.

The Dnmt1 methyltransferase contains a C-terminal region of ~500 amino acids that is related to bacterial type II cytosine-5 methyltransferases (Lauster et al., 1989; Posfai et al., 1989). A reverse genetics approach has identified a family of mammalian DNA (cytosine-5) methyltransferases by using the type II bacterial enzymes to search the expressed sequence tag (EST) databases (Okano et al., 1998a). These proteins, Dnmt2 and Dnmt3 contain the highly conserved cytosine-5

methyltransferase motifs however the Dnmt2 enzyme did not methylate DNA in vitro or in vivo (Okano et al., 1998b). Furthermore, targeted deletion of *Dnmt2* in ES cells did not affect growth or morphology suggesting Dnmt2 is not essential in contrast to Dnmt1 (see section 1.3.1). Dnmt2 is homologous to the fission yeast pmt1 (*p*ombe *m*ethyl*t*ransferase1) which does not possess any detectable methyltransferase activity (Wilkinson et al., 1995). The Dnmt3 protein showed only a modest ability to enzymatically modify unmethylated and hemi-methylated DNAs. Targeted deletion by homologous recombination will clarify the role of *Dnmt3* in mouse development and de novo methylation.

### **1.1.3 Intrinsic mutability of 5-methylcytosine**

The methyl group present in m<sup>5</sup>C can directly affect protein-DNA interactions (Iguchi-Ariga and Schaffer, 1989; Bednarik et al., 1991), increase the thermal stability of C-G base pairs by several degrees (Collins and Myers, 1987) and alter the energetics of structural transitions from the B to Z forms of DNA (Behe and Felsenfeld., 1981). The base m<sup>5</sup>C also possesses an intrinsic mutability. Cytosine moieties in DNA can spontaneously deaminate to uracil and dedicated uracil DNA glycosylases repair mismatches arising from C to U transitions. Deamination of m<sup>5</sup>C however generates thymine which cannot be repaired by the uracil DNA glycosylase pathway. Hydrolytic deamination of m<sup>5</sup>C allows a single-step conversion between nucleotides that is not possible for the other bases within DNA (Lindahl and Nyberg, 1974). Human cells contain an activity that recognises T/G (T DNA glycosylase,

TDG) in m<sup>5</sup>CpG/TpG mismatches and preferentially removes the T (Nedderman and Jiricny, 1993). However, m<sup>5</sup>CpG to TpG mutations account for almost 50% of mutations in colon cancer (Hollstein et al., 1996) and approximately one-third of single base pair mutations causing human genetic disease occur in m<sup>5</sup>CpG dinucleotides (Cooper and Youssoufian, 1988). As a result, the dinucleotide CpG is depleted in the vertebrate genome and this is likely the result of the loss of m<sup>5</sup>CpG to its deamination products, TpG and CpA (Bird, 1980). Lieb and Rehmat (1997) have reported a markedly reduced rate of spontaneous deamination of m<sup>5</sup>C in non-dividing bacteria which was similar to m<sup>5</sup>C in solution. The authors suggest that the mutation hot spots at m<sup>5</sup>C in rapidly dividing cells is the result of insufficient time for T/G mismatch repair pathways to operate before the next round of replication. Interestingly, the mutation frequency of p53 in tissues that are growing rapidly is increased relative to the mutation rate observed in slowly growing tissues (Hollstein et al., 1996).

## **1.2 DNA methylation outside the vertebrate lineage**

### **1.2.1 Plant DNA methylation**

The genomic DNA of plants contains m<sup>5</sup>C at the symmetric sites, CpG and CNG and as much as 30% of total cytosine residues are modified to m<sup>5</sup>C (Gruenbaum et al., 1981b). Plants also contain a constitutively non-methylated CpG-rich fraction that co-localises with the 5' ends of genes and are analogous to vertebrate CpG islands (Antequera and Bird, 1988). Transcriptional silencing of transgenes in plants correlates with their hypermethylation (Ye and Singer, 1996) and gene silencing in

plants usually arises following sequence duplication events occurring either *in cis* or *in trans* in a phenomenon termed homology dependent gene silencing. In active transgenes there is a low basal level of methylation (~10% of cytosines) and this increases to 80-90% upon inactivation. DNA (cytosine-5) methyltransferases in plants act preferentially on hemi-methylated DNA and have a high specificity for CpG or CAG and CTG (Finnegan and Dennis, 1993; Pradhan and Adams, 1995). In a forward genetic screen in *Arabidopsis*, two genes *DDM1* and *DDM2* (Decrease in DNA Methylation) were identified that are required to maintain normal levels of m<sup>5</sup>C (Vongs et al., 1993). In *DDM1* mutants a 70% reduction in genomic m<sup>5</sup>C levels was observed, with hypomethylation of repeated sequences occurring initially. Interestingly, *DDM1* encodes a protein with a high degree of similarity to the SWI2/SNF2 suggesting that the DDM1 protein alters genomic methylation patterns by influencing chromatin structure (Jeddeloh et al., 1999). DDM1 may simply operate as a co-activator for the expression of a methyltransferase or alternatively a more interesting hypothesis is that DDM1 may operate to remodel nucleosomes that then increases the accessibility of DNA methyltransferases to hemi-methylated targets in nascent chromatin.

### **1.2.2 Fungal DNA methylation**

DNA methylation has not been detected in the model yeasts *Saccharomyces cerevisiae* (Proffitt et al., 1984) and *Schizosaccharomyces pombe* (Antequera et al., 1984). In *Ascobolus* however, DNA repeats are densely methylated and inactivated

(Goyon and Faugeron, 1989). This silencing phenomenon is similar to homology dependent gene silencing in plants where duplicated elements are de novo methylated and inactivated. This process is termed “methylation induced premeiotically” (MIP) (Goyon et al., 1994). The inactivation process occurs after fertilisation in dikaryotic tissue and the methylation is subsequently maintained. *Masc1* (Mtase from Ascobolus 1) is required for this de novo methylation and is homologous to bacterial type II and eukaryotic maintenance type DNA (cytosine-5) methyltransferases however *Masc1* does not have demonstrable de novo methyltransferase activity in vitro (Malagnac et al., 1997). Disruption of *masc1* in homokaryons has no phenotype and the gene is thus not essential for vegetative growth. Furthermore, homokaryons were able to perform maintenance methylation of a previously inactivated transgene however crossing of *masc1* mutants results in sterility in dikaryons and progeny are arrested at an early stage of sexual reproduction. No vertebrate homologs of *masc1* have been identified to date.

### **1.2.3 Invertebrate DNA methylation**

Vertebrate genomes are characterised by a global low density of methylation as are the earliest examples of vertebrates, the jawless fishes. In contrast, the genomes of certain chordates such as *Ciona* and *Amphioxus* have intermediate levels of methylation that are present in discrete compartments or appear to lack detectable methylation genome as is the case for the fruitfly, *Drosophila melanogaster* (Rae and Steele, 1979). Initially it was thought these methylated domains were present

exclusively at repetitive sequences with genes located in the non-methylated compartments (Bird et al., 1979) however subsequent findings suggest that significant numbers of genes in *Ciona* and *Amphioxus* are located within both the methylated and unmethylated regions (Tweedie et al., 1997; Simmen et al., 1999). It is unclear what directs certain regions for modification in the fractionally methylated genomes of these organisms but the establishment of the methylation pattern clearly does not operate in a analogous manner to that seen with the de novo methylation of transgenes in homology dependent gene silencing in plants or the similar MIP phenomenon in *Ascobolus*, as genes and repetitive elements are found in both fractions (Tweedie et al., 1997). In contrast the genome of the invertebrate *Physarum* is methylated and the retrotransposon-like element, Tp1 is located predominantly within the methylated fraction (Rothnie et al., 1991).

Simpson et al., (1986) were unable to detect m<sup>5</sup>C in the genome of *C. elegans* at any stage of development. A qualitative analysis using the methylation-sensitive and methylation-insensitive isoschizomers, *HpaII* and *MspI* did not detect any differences in the restriction pattern following digestion with these enzymes and no methylation of a repeated element (Tc1) was detectable by Southern analysis. The use of the *HpaII* and *MspI* isoschizomers to detect DNA methylation in this study biased the search for genomic methylation as modification outwith their symmetric target sequence (CCGG) would not have been detected. The sensitive and quantitative technique of high performance liquid chromatography was also applied



to *C. elegans* genomic DNA and the authors concluded that fewer than 1:10000 cytosines (<0.01 mole percent) were methylated in any sequence context. Therefore, the possibilities for m<sup>5</sup>C in the regulation of transcription and mediating other biological effects are limited in *C. elegans*. In eukaryotic organisms, there are three other covalently modified nucleotide bases identified in DNA in addition to m<sup>5</sup>C: N6-methyladenine, hydroxymethyluracil and  $\beta$ -D-glucosyl-hydroxymethyluracil. Many restriction enzymes are inhibited by the presence of modified bases within their recognition sequences (Berkner et al., 1979; Huang et al., 1982) and covalent modifications of DNA are often inferred from unusual digestion patterns of genomic DNA when compared with the same cloned DNA sequence (Bernards et al., 1984; Pays et al., 1984). Despite the fact that there are no reported cases of anomalous resistance of genomic DNA to restriction endonucleases in the nematode worm, novel forms of covalent modification in *C. elegans* cannot be excluded.

### **1.3 DNA methylation and development**

The genomes of the model genetic organisms *Caenorhabditis elegans* and *Drosophila melanogaster* lack detectable DNA methylation (Simpson et al., 1986; Rae and Steele, 1979) and this has lead to questions being raised as to the functional consequences and the relevance of this modification in higher metazoans. DNA methylation has long been correlated with transcriptional silencing and this has been demonstrated on numerous occasions in vitro by methylation of gene constructs and transfection into cell-lines (Vardimon et al., 1982; Stein et al., 1982). In addition, the nucleoside analogue 5-aza-2'-deoxycytidine (5-azaC) which alters genomic

methylation patterns by covalently trapping the maintenance methyltransferase, Dnmt1 (Juttermann et al., 1994), was found to promote aberrant gene expression (Jones and Taylor, 1980).

### **1.3.1 Reduced genomic methylation and developmental abnormalities**

Gene targeting experiments that introduced mutations at the murine DNA (cytosine-5) methyltransferase (MTase) locus (*Dnmt1*) were found to have an embryonic lethal phenotype with the mice dying early in development (Li et al., 1992; Lei et al., 1996). Assuming that the lethality arises as a result of reduced genomic methylation and not some unknown function of the Dnmt1 protein, these findings suggest that the completion of embryonic development in higher vertebrates requires normal patterns of methylation. Embryonic stem (ES) cells homozygous for this mutation in *Dnmt1* grown in the presence of differentiation-inhibitory activity were found to be viable in culture but died rapidly following differentiation (Li et al., 1992). This has led to the suggestion that DNA methylation is dispensable for pluripotent embryonic stem cells and is only required for differentiated somatic cells (Jaenisch, 1997). Subsequently, wild type genomic methylation patterns have been found to be a necessary prerequisite for development in a more diverse group of organisms following the finding that mutations that reduce levels of m<sup>5</sup>C in the DNA of a vascular plant (Vongs et al., 1993) and fungi (Foss et al., 1993; Malagnac et al., 1997) cause defects in chromosome stability and segregation. Furthermore, a partial demethylation phenotype has been generated by inhibition of the *Arabidopsis*

methytransferase and this was found to result in developmental abnormalities following expression of the antisense RNA (Ronemus et al., 1996). Taken together, these findings suggest an important role for DNA methylation in the development of both higher plants and mammals.

### **1.3.2 Phenotype of the *Dnmt1* knockout mouse**

Why do organisms so diverse require normal genomic methylation patterns for the completion of development ? A potential explanation may lie in the genome defence model (Bestor, 1990; Yoder et al., 1997) which posits that the main role of DNA methylation in animals is to restrain the spread of endogenous elements, proviruses and other selfish sequences in germ cells. The developmental abnormalities seen in these organisms might thus arise through perturbation of gene expression as a result of increased transposition into active areas of chromatin. However, determining whether methylation is targeted preferentially to selfish elements as opposed to genes within a mammalian genome is difficult because of the low-level genome-wide methylation seen in higher vertebrates. Simmen et al., (1999) have sought to test this prediction of the genome defence model - a co-linearity between DNA methylation and known selfish elements - by analysing the specificity of methylation patterns in the genome of the invertebrate urochordate, *Ciona intestinalis*. In contrast to the global modification by methylation evident in vertebrate DNA, *Ciona* possesses a fractionally methylated genome (Tweedie et al., 1997). Thus any bias in predicted gene or transposon location would be informative in this organism. Strikingly, multiple copies of transposons were found to reside within unmethylated regions

whereas genes were located predominantly in methylated domains, a result not easy to reconcile with the genome defence model. A caveat of these findings however is that it is unknown whether methylation regulates gene activity in invertebrates and methylation may simply operate in a mechanistically different manner than in vertebrates. An alternative explanation for the phenotype of the *Dnmt1* knockout mouse and the developmental abnormalities seen in a range of organisms with reduced levels of genomic m<sup>5</sup>C is that DNA methylation is responsible for developmental gene control. Walsh and Bestor (1999) were unable to find ectopic expression of genes reported to be regulated in a methylation-dependent manner in the *Dnmt1*<sup>-/-</sup> mutant embryos. However, Li et al., (1993) found a failure in *Dnmt1*<sup>-/-</sup> ES cells to regulate expression of imprinted genes. In the mutant embryos, the three imprinted genes *H19*, *Igf-2* and *Igf-2r* were abnormally expressed. The normally repressed paternal allele of *H19* was active whereas in contrast the normally active paternal *Igf-2* and maternal *Igf-2r* were repressed. In addition, ectopic expression of the *Xist* transcript in differentiated *Dnmt1*<sup>-/-</sup> ES cells was observed from the normally silenced inactive X-chromosome which correlated with a dramatic increase in apoptosis (Panning and Jaenisch , 1996). These findings suggest the embryonic lethality observed in the *Dnmt1* knockout mice might arise because of aberrant gene dosage of X-linked and imprinted loci as a result of a failure in methylation-mediated repression.

## **1.4 Molecular mechanisms for methylation-mediated repression**

### **1.4.1 Direct inhibition of transcription factor binding**

How might the reduction in levels of genomic methylation in the *Dnmt1*<sup>-/-</sup> alter transcriptional regulation of imprinted and sex-linked genes? The simplest conceivable model for methylation-mediated transcriptional repression is by direct interference of transcription factor binding. Indeed, several well understood transactivators are unable to bind their target sites when methylated. Examples include the cyclic AMP response element binding protein (CREB; Iguchi-Ariga and Schaffer, 1989) as well as the ubiquitous transcription factor, NF-κB (Bednarik et al., 1991). This model cannot however explain silencing, attributed to methylation, from promoters where no identified methylation-sensitive transcription factors are present. Indeed, some factors such as Sp1 are known to be capable of binding irrespective of the methylation status of their recognition sequence and relatively few transcription factors contain the dinucleotide CpG within their cognate sites. These findings argue against a widespread role for direct inhibition of transcription factor binding in mediating the repression associated with DNA methylation.

### **1.4.2 Indirect inhibition of transcription factor binding**

An indirect model for transcriptional inhibition posits that nuclear proteins with affinity for methyl-CpG sites bind within promoter regions and thereby preclude the assembly of the transcriptional machinery. Evidence for an indirect mechanism of methylation-mediated transcriptional repression comes from a number of reports. Murray and Grosveld (1987) demonstrated that the transcriptional inhibition of the

$\gamma$ -globin gene was not dependent on specific sites within the promoter, since the presence of methyl-CpGs in different locations within the promoter was sufficient to repress transcription. The authors concluded that a minimal methylation free zone was required for expression, and hypothesised that proteins with affinity for methyl-CpGs were responsible for repression. Involvement of chromatin in mediating the effects of DNA methylation was suggested by experiments from Buschhausen et al. (1985, 1987). Buschhausen et al., (1987) observed that methylation of the herpes simplex thymidine kinase (HSV-tk) gene and subsequent microinjection into mouse L929 cell nuclei did not lead to immediate transcriptional inhibition. Indeed, the HSV-tk gene remained transcriptionally active for ~8 hours following injection into cells before being repressed. This observation is incompatible with a model where methylation of CpG would directly interfere with assembly of the initiation complex and implied that time was required for interactions with nuclear components to abrogate transcription. In addition, Keshet et al. (1986) found that M13 gene constructs when methylated to completion and transfected into mouse L929 cells were preferentially assembled into nuclease resistant chromatin. This was in contrast to unmethylated sequences which integrated in a potentially active DNaseI-sensitive conformation. These results suggested that methylation may exert its effects on gene transcription by altering both specific and non-specific interactions between DNA and nuclear proteins. Furthermore, it is known that methyl-CpG sites are protected from cleavage by non-specific nucleases (Solage and Cedar, 1978) and against restriction enzymes that specifically recognise CpGs (Antequera et al., 1989). A

major advance in the understanding of DNA methylation was the establishment of a link between this phenomenon and that of the other major epigenetic modification of chromatin - histone acetylation.

## **1.5 Histone acetylation and deacetylation in gene control**

### **1.5.1 Introduction**

Eukaryotic DNA is packaged into chromatin which thus affects every cellular process that requires access to the DNA template. The nucleosome is the primary unit of chromatin organisation and is composed of two molecules each of the core histones H2A, H2B, H3 and H4 with approximately two turns of DNA wrapped around this octamer of proteins. The assembly of DNA into nucleosomal chromatin represses transcription in vivo and in vitro and the structure of chromatin and the regulation of gene expression are as a result intimately linked processes. DNA in chromatin thus serves a dual role in both the packaging of DNA and globally repressing gene transcription. The amino termini of the core histones extend from the nucleosome core and are subject to post-translational modifications by ADP-ribosylation, methylation, phosphorylation and acetylation. Since the identification of core histone acetylation (Allfrey et al., 1964) a general correlation has been found to exist between the acetylation status of a chromosomal domain and transcriptional activity.

### **1.5.2 Histone acetylation and gene activation**

Hyperacetylated histones are found preferentially within active, euchromatic domains (Hebbes et al., 1994) whereas hypoacetylated histones are found to reside within areas that are silent and heterochromatic (Braunstein et al., 1993). Euchromatin decondenses during the interphase of the cell cycle and is gene-rich whereas conversely heterochromatin condenses during interphase and is gene-poor. The amount of histone acetylation is the result of an equilibrium between opposing histone acetyltransferase and histone deacetylase activities. Acetylation of the nucleosomal core histones occurs on the highly conserved N-terminal lysine residues that lie outside the nucleosome. The N-terminal tails are known to play a key role in denying transcription factors access to the nucleosomal DNA templates (Lee et al., 1993) as well as interacting with regulatory proteins (Johnson et al., 1990). The hyperacetylation of histone tails is thought to direct an allosteric destabilisation of the nucleosome by neutralising the electrostatic interaction between the positively charged lysine residues and the negatively charged phosphodiester backbone of the DNA template. Given that histone acetylation is conserved throughout eukaryotes, it is an attractive central point to regulate gene activity and the enzymatic activities that establish these dynamic patterns on histones have been intensively studied.



### 1.5.3 Histone acetyltransferases

Following the purification and sequencing of the gene that encodes the *Tetrahymena* p55 nuclear histone acetyltransferase, a surprising homology was identified to a previously isolated yeast transcriptional regulator GCN5. Brownell et al., (1996) subsequently demonstrated that GCN5 possessed acetyltransferase activity and a central region of GCN5 showed homology to previously identified histone acetyltransferases (Kleff et al., 1995). The *S. cerevisiae* heterotrimeric GCN5/ADA2/ADA3 activator complex was identified by Guarente and colleagues screening for mutations in genes that confer resistance to the potent and toxic chimaeric activator GAL4-VP16 (Berger et al., 1992). The genes identified from the screen were reasoned to encode proteins necessary for VP16 activation of transcription. Subsequently, ADA2 was proposed to serve as an adaptor linking activation domains and the basal transcriptional machinery and ADA2 has been shown to interact with the TATA-binding protein (TBP) of TFIID as well as with generic acidic activation domains (Barlev et al., 1995; Silverman et al., 1994). This suggested that GCN5 might be targeted to promoters to activate gene transcription by ADA2 associating with sequence-specific DNA binding proteins. Mutations in *GCN5* diminished the ability of the DNA-binding transcription factor GCN4 to activate transcription (Georgakopoulos and Thireos, 1992) suggesting the potential for GCN5 to be recruited to a gene through interactions with transcription factor(s) to modulate transcription. *GCN5* was found to be non-essential for yeast viability which raised the possibility that multiple redundant and overlapping genes might

exist encoding acetyltransferases (Georgakopoulos and Thireos, 1992). This suggestion was confirmed by the identification of the p300/CBP-associated factor (PCAF) that was capable of acetylating the nucleosomal core histones (Yang et al., 1996a). The p300/CBP (CREB-binding protein) itself was found to also possess histone acetyltransferase potential (Ogryzko et al., 1996) and is known to mediate gene regulation by a diverse array of DNA-binding transcription factors (Kamei et al., 1996). TFIID, the transcription factor complex of the TATA-binding protein (TBP) and the TBP-associated factors (TAFs) contains histone acetyltransferase activity and TAF<sub>II</sub>250 was found to possess intrinsic histone acetyltransferase potential (Mizzen et al., 1996). This suggests that the basal transcriptional machinery might destabilise nucleosomes over promoters to activate gene transcription. PCAF and GCN5 are related proteins but no significant sequence similarity or structural homology is thought to exist between these proteins and the other identified histone acetyltransferases. PCAF is able to acetylate histone H3 in a nucleosome while human GCN5 can acetylate both free and nucleosomal histones H3 and H4 (Xu et al., 1998). In contrast, recombinant p300/CBP can acetylate all core histones in a nucleosomal context and acetylation occurs on histone H4 at K5, 8, 12 and 16 which are the same positions modified *in vivo* (Yang et al., 1996). A number of major questions remain as to whether histone hyperacetylation is a cause or effect of the chromatin disruption process at activated gene promoters and how the targeting of chromosomal regions for accumulation of hypo- or hyperacetylated histones (or other non-histone targets) occurs.

The bromodomain motif is widely conserved during evolution and is commonly found in transcription factors implicated in transcriptional regulation at the level of the chromatin template (Haynes et al., 1992; Jeanmougin et al., 1997). The bromodomain is present in the vast majority of identified histone acetyltransferases including the yeast and human GCN5s (Brownell et al., 1996; Georgakopoulos and Thireos, 1992), PCAF (Yang et al., 1996), CBP/p300 (Bannister and Kouzarides, 1996) and human TAF<sub>II</sub>250 and *Drosophila* TAF<sub>II</sub>230 (Mizzen et al., 1996; Ruppert et al., 1993) proteins. The bromodomain is not required for histone acetyltransferase activity (Candau et al., 1997) and this suggests it might serve some ancillary function for regulating these enzymatic activities or targeting to chromatin sites. Secondary structure predictions suggest that the bromodomain is composed of four hydrophobic  $\alpha$ -helices (Jeanmougin et al., 1997) that might constitute a protein-protein interaction motif. Bromodomains are frequently found in association with other conserved motifs that are known to serve a catalytic function. In addition to its presence in histone acetyltransferases, the bromodomain is interestingly also found in the yeast transcriptional regulator SWI2/SNF2, a component of the large SWI/SNF complex (Peterson et al., 1994; Cairns et al., 1994). Genetic suppressors of SWI2/SNF2 encode both histone and non-histone proteins, suggesting the SWI/SNF complex counteracts histone-mediated and non-histone based transcriptional repression (Kruger et al., 1995). SWI2/SNF2 also shares homology including a bromodomain with brahma (*brm*) an activator of the homeotic gene clusters in *Drosophila*.

### 1.5.4 Histone deacetylases

The recruitment by sequence-specific DNA-binding factors of transcriptional regulators with intrinsic enzymatic ability to modulate the architecture of chromatin to control gene regulation has become an attractive way to model how the structural core histones interact with the transcriptional machinery. This model while first supported by studies on transcriptional activators/co-activators was subsequently widened to include transcriptional repressors/co-repressors as it became clear that the repression of gene transcription utilised similar yet opposing mechanisms to that used to activate gene transcription.

Employing an affinity matrix with the ligand trapoxin, a high affinity irreversible histone deacetylase inhibitor, the HDAC1 protein was identified in complex with the WD40 repeat protein RbAp48 (Taunton et al., 1996). The HDAC1 protein showed 60% sequence identity to *S. cerevisiae* RPD3 that was previously found to be required for full repression and activation of a subset of genes (Vidal and Gaber, 1991). In *Drosophila* mutation of an *RPD3* homolog resulted in an increase in position effect variegation suggesting an increase in silencing (De Rubertis et al., 1996). Deletions in *RPD3* in yeast resulted in the predicted hyperacetylation of histones H3 and H4 however somewhat paradoxically resulted in an increase in repression at telomeres (Rundlett et al., 1996).

The bifunctional transcriptional regulator YY1 is a zinc-finger transcription factor that is thought to be involved in the regulation of cell proliferation and differentiation

(Shi et al., 1991; Shrivastava and Calame, 1994). Yang et al., (1996b) demonstrated interaction between YY1 and the mammalian RPD3 co-repressor. Fusion of RPD3 to a heterologous DNA binding domain could direct a dose dependent transcriptional inhibition of a cis-linked reporter gene. Furthermore site direct mutagenesis of a glycine rich motif in YY1, necessary for interaction with RPD3 abrogated the ability of YY1 to repress transcription. This suggested that the tethering of RPD3 to YY1 bound on DNA could negatively regulate gene transcription by removal of acetyl moieties on the N-terminal lysine residues of the nucleosomal core histones.

A number of transcriptional repressors have now been shown to associate with histone deacetylases and one of the best characterised is the transcriptional repressor, Mad. Mad can form heterocomplexes with the DNA targeting protein, Max which binds E-boxes on promoters (Ayer et al., 1995; Alland et al., 1997). The Mad-Max heterodimer can influence cell growth and differentiation in mammalian cells (Ayer et al., 1995). The DNA-binding Max protein can heterodimerise with the Myc family of bHLH-ZIP proteins that include Myc, Mad and Mxi-1 (Ayer et al., 1993) in order to influence the activity of downstream target genes. The Myc-Max complex bound at E-boxes is an activator of transcription and is the complex found in proliferating cells whereas the Mad-Max transcriptional repressor complex is the form found in differentiated cells and acts in a manner antagonistic to the Myc-Max complex. Two mammalian proteins, mSin3A and mSin3B have been identified that interact with the Mad repressor and are homologous to yeast Sin3. Mutation of the second of four

paired amphipathic helix (PAH) domains of mSin3A abolishes binding to the Mad-Max complex and prevents the complex from repressing transcription (Ayer et al., 1995). Three different histone deacetylases (HDACs), HDAC1-3 have been identified in mammalian cells (Grunstein et al., 1997) and appear to possess overlapping activities. Co-immunoprecipitation revealed the Mad, Sin3 proteins and the histone deacetylases to reside in a complex together (Alland et al., 1997). In experiments employing the histone deacetylase inhibitor trichostatin A (TSA) (Yoshida et al., 1990), the ability of the Mad-Max complex to repress transcription was attenuated demonstrating a clear role for deacetylation in gene control. The mechanism by which unliganded nuclear hormone receptors represses transcription also utilises the mSin3/HDAC pathway (Alland et al., 1997; Heinzl et al., 1997). Unliganded thyroid hormone and retinoic acid receptors bind the transcriptional co-repressor, NCoR (Horlein et al., 1995) which interacts with the mSin3/HDACs (Alland et al., 1997; Heinzl et al., 1997) to repress transcription and this repression is alleviated entirely by inhibition of histone deacetylases using TSA (Wolffe et al., 1997). Furthermore, repression of meiotic genes by the yeast Ume6 also involves association with a Sin3/Rpd3 complex. These findings taken together suggest that the modification of the nucleosomal core histones by acetyltransferases and deacetylases is a widely conserved mechanism to modulate eukaryotic gene transcription.

## **1.6 Methyl-CpG binding Proteins (MeCPs)**

Methylation of mammalian DNA can lead to repression of transcription and alteration of chromatin structures. Both effects are potentially the result of interactions between methylated sites in the genome and proteins with affinity for methylated DNA. A number of activities have been identified in mammalian nuclei which associate with methylated DNA irrespective of sequences flanking the methyl-CpG site and represent strong candidates for mediating the global consequences of methylation. This relaxed sequence specificity distinguishes these activities from human methylated DNA-binding protein 1 (MDBP1; Wang et al., 1986) and chicken methylated DNA-binding protein 2 (MDBP2; Pawlak et al., 1991) each of which require specific nucleotide sequence contexts to bind to methylated sites.

Methyl-CpG binding protein 1 (MeCP1) was identified originally as an activity present in mouse nuclear extracts that bound to methylated DNA containing a high density of methyl-CpGs in a variety of sequence contexts (Meehan et al., 1989). MeCP1 has been implicated as a mediator of methylation-associated transcriptional silencing (Boyes and Bird, 1991, 1992). The activity was detectable in a range of somatic tissues but was conspicuously absent from cell lines derived from embryonic lineages suggesting that the protein was dispensable for growth in undifferentiated cells. MeCP2 was first detected during attempts to obtain MeCPs by expression cloning (Lewis et al., 1992). MeCP2 is an abundant nuclear protein that can bind a single methyl-CpG pair in contrast to MeCP1. Immunofluorescence using an anti-

MeCP2 antibody showed a preferential localisation to the pericentromeric heterochromatin in the mouse genome which closely matched the mouse major satellite DNA. The consensus sequence of the mouse major satellite is a 234bp AT-rich unit which contains eight CpGs and at least seven have been shown to be methylated (Horz and Altenburger, 1981; Manuelidis, 1981). This satellite therefore contains >40% of all the methyl-CpGs in the mouse nucleus. Miller et al., (1974) directly visualised this non-random distribution in the m<sup>5</sup>C distribution by anti-m<sup>5</sup>C antibody and the staining pattern closely matches that with the anti-MeCP2 antibody.

MeCP2, by virtue of its ability to bind to a single methyl-CpG pair, could explain transcriptional repression of weak promoters brought about by the presence of a small number of methyl-CpG moieties. The high abundance of MeCP2 (~10<sup>6</sup> molecules per nucleus) may explain the genome-wide protection of CpG sites from nuclease digestion (Solage and Cedar, 1978; Antequera et al., 1989). MeCP2 is known to be abundant in brain tissue where methyl-CpG sites are found to be particularly resistant to the action of exogenous nucleases. In contrast, embryonic PC13 cells which lack both MeCPs, exhibit markedly reduced levels of protection (Antequera et al., 1989).



### **1.6.1 The methyl CpG-binding domain (MBD)**

Deletion analysis determined that binding to methylated DNA required a minimal region between amino acids 89 and 162 at the N-terminus of MeCP2 (Nan et al., 1993). This peptide was named the methyl CpG-binding domain (MBD) and this region is known to be invariant in rat, mouse, chicken and human MeCP2s. DNA methylation has been causally implicated in the localisation of MeCP2 in chromatin (Nan et al., 1996). The preferential localisation of MeCP2 in heterochromatic regions suggested by immunofluorescence was addressed by transfecting expression constructs for MeCP2- $\beta$ geo fusions into a mouse ES cell line that show a reduced level of genomic m<sup>5</sup>C (Li et al., 1992). The observed disruption in the ability of localise MeCP2 to heterochromatin in *Dnmt1*<sup>-/-</sup> cells compared to wild type ES cells indicates that DNA methylation is a major determinant of the sub-cellular localisation of MeCP2 in vivo.

### **1.6.2 MeCP2 is a transcriptional repressor and is essential for mouse development**

Nan et al., (1997) found that native and recombinant MeCP2 were able to repress transcription from a reporter template in a methylation-specific fashion. An internal deletion of the N-terminal MBD of MeCP2 abrogated the ability to silence transcription in vitro in primer extension assays from a methylated template. A GAL4-MeCP2 fusion was also found to be capable of repressing transcription of a reporter gene in vivo. This silencing effect was mapped to a highly basic central region of the protein, the transcriptional repression (TR) domain. Earlier experiments

by Hug et al., (1995) had suggested that promoter methylation *per se* was not always required for transcriptional inhibition but that distally situated sites of high density methylation could operate in cis to silence linked reporters. MeCP2 was found to be capable of repressing transcription at a distance in vitro and in vivo. MeCP2 can interact with methylated sites in the context of chromatin and intriguingly was capable of displacing histone H1. This replacement of histone H1 by MeCP2 may explain the distinct chromatin structure that occurs within methylated domains of chromatin. Histone H1 has been shown to be an abundant component of heterochromatin (Weintraub, 1984) and to be associated with nucleosomes containing m<sup>5</sup>C (Ball et al., 1983). In contrast, H1 is absent from CpG island chromatin (Tazi and Bird, 1990) which is characteristic of active, autosomal housekeeping genes, and is depleted in chromatin which contains active genes (Bresnick et al., 1991). Histone H1 is known to bind the internucleosomal linker, to play a role in higher-order stabilisation of chromatin structure (Thoma et al., 1979) and to be capable of transcriptional silencing in vitro (Wolffe, 1989; Croston et al., 1991). Previous reports describing interactions between H1 and methylated DNA demonstrating transcriptional repression have been done with naked DNA (Levine et al., 1993; Johnson et al., 1995; McArthur and Thomas, 1996) and it is assumed that the effects observed reflect in vivo interactions of H1 and chromatin. This assumption is made in the light of the fact that the interaction of H1 in chromatin is primarily with DNA rather than with the core histones (Thoma et al., 1979). Campoy et al. (1995),

however found no affinity of H1 for methylated DNA with chromatin assembled on plasmids in a *Xenopus* oocyte extract in the presence of added rat histone H1.

MeCP2 has been mapped to the X-chromosome (Quaderi et al., 1994). To assess the functional significance of the protein in vivo, hemizygotes were generated by targeted disruption of the locus in ES cells employing a promoterless *lacZ* construct (Tate et al., 1996). The mutant ES cells lacking MeCP2 grew normally during in vitro passage and were capable of considerable differentiation. Chimaeric embryos however exhibited developmental defects proportional to the contribution of mutant ES cells. Embryos with the highest levels of chimaerism failed to gastrulate leading to embryonic death at mid-gestation. Similarities exist between the *MeCP2* and the *Dnmt1* mutants. Both genes apparently are superfluous to survival of undifferentiated cells yet are absolutely required during development.

## **1.7 The MBD protein family**

### **1.7.1 Comparative genomics and expressed sequence tags (ESTs)**

The development of large scale sequencing projects has lead to an exponential increase in the amount of sequence information available to investigators, resulting in the rapid identification of new genes and their corresponding cDNAs from a diverse array of organisms. In particular, the development of expressed sequence tag (EST) based libraries, such as dbEST (Boguski et al., 1993) and XREFdb (Bassett et al., 1997), for the model organisms *Drosophila melanogaster*, *Caenorhabditis elegans* and the laboratory mouse, *Mus musculus* has proven to be invaluable tools for the

isolation of novel cDNAs related to conserved functional domains of interest. In parallel with the growth of these large scale sequencing projects, the fields of comparative and functional genomics have developed and represent a new multi-organismal approach to biology that have lead to major advances in genetic analysis. However, with these vast amounts of information to hand, a requirement for a matching scale-up and new approaches to functional studies is required in order to assign function to these new sequences. To this end, a consortium has been formed to disrupt all *C. elegans* genes and strategies are being developed in the mouse to rapidly generate mutants to analyse gene function (Zambrowicz et al., 1998). ESTs are random, single-pass unedited cDNAs and have become powerful molecular tools for novel gene identification (Cross et al., 1997), candidate disease gene identification (Scharf et al., 1998), genomic sequence interpretation (Jiang and Jacob, 1998), the development of microarrays for functional genomics (Schena et al., 1998), comparative genomics (Makalowski et al., 1996) and the development of gene based maps of entire genomes (Schuler et al., 1996). The EST based libraries of sequence information have gradually began to usurp the more traditional approaches of novel gene identification such as degenerate RT-PCR based applications and low stringency library screening with functional domains of interest. These reverse genetic approaches, where defined motifs are employed to isolate additional proteins through database screens has been employed successfully to identify the protein mediators of the biological consequences of the methylation of DNA in higher vertebrates.

### **1.7.2 Employing ESTs to identify novel MBD-containing proteins**

Methylation can inhibit transcription factor binding directly (Watt and Molloy, 1988; Iguchi-Arigo and Schaffer, 1989), but methylated sequences are not efficiently silenced in cell lines where MeCP1 and MeCP2 are undetectable (Boyes and Bird, 1991; Levine et al., 1991) suggesting indirect mechanisms are more important. It is plausible that the transcriptional suppression, chromatin alterations, late replication and inhibition of recombination known to correlate with DNA methylation are the biological consequences of association between MeCPs and methyl-CpG sites in the genome. Thus the identification of novel MeCPs is an area of intense interest, as these proteins may well serve as the molecular integrators of diverse biological phenomena. MeCP2 like many transcription factors exhibits a modular organisation, containing an N-terminal methyl-CpG binding domain (MBD) (Nan et al., 1993; Nan et al., 1996) and a centrally located transcriptional repression domain (TRD) (Nan et al., 1997). An N-terminal region of 80 amino acids of MeCP2 has been delineated as the minimal region necessary to confer specific association with methylated DNA (Nan et al., 1993) and an intact MBD is absolutely required for the association of MeCP2 and chromatin (Nan et al., 1996). Whilst no TRD-containing proteins, apart from MeCP2 homologs, have been identified within the EST databases, the MBD of MeCP2 has been found to be widely conserved through many lineages.

In order to identify new sequences with homology to the MBD, Cross et al., (1997) searched the EST databases with the 80 amino acid MBD of MeCP2 reasoning that

this motif would be conserved in other methyl-CpG binding proteins. This hypothesis was substantiated by the identification of a number of ESTs that exhibited strong similarity to the MBD of MeCP2. A human retinal cDNA library was found to contain two ESTs that exhibited 43.6% identity and 52.7% similarity over 55 amino acids comprising the MBD. This protein is termed MBD1 (Hendrich and Bird, 1998). Recombinant MBD1 was found to bind symmetrically methylated DNA specifically and this property was dependent on an intact MBD. Moreover, MBD1 was demonstrated to be capable of repressing transcription, as addition of recombinant protein to a series of in vitro primer extension reactions repressed transcription in a methylation-dependent manner. Hendrich and Bird, (1998) have extended this EST based approach to identify further MBD-containing cDNAs and have characterised three additional proteins (MBD2, MBD3 and MBD4) expressed widely in mouse and human tissues. Using a combination of both in vitro and in vivo assays, MBD1, MBD2 and MBD4 were demonstrated to complex specifically with symmetrically methylated DNA irrespective of sequence context. Thus, with MeCP2 these methylated DNA-binding proteins represent strong candidates for operating as the protein mediators for the biological influence of DNA methylation.

## **1.8 DNA methylation and histone deacetylation**

The recent advances in the characterisation of the repressor/co-repressor complexes that mediate transcriptional inhibition by recruitment of complexes that deacetylate the core histones H3 and H4 have aided in the elucidation of the molecular mechanism

by which DNA methylation silences gene transcription. Hsieh (1994) reported that following transfection of a methylated reporter construct and treatment with the histone deacetylase inhibitor sodium butyrate, derepression of the normally silenced state was observed suggesting the involvement of histone deacetylation in mediating the transcriptional silencing by CpG methylation. Selker (1998) also observed a derepression of previously silenced and methylated genes in the *Neurospora* system following treatment with trichostatin A, implicating DNA methylation in the control histone acetylation. Similar observations were made by Chen and Pikaard (1997) studying the phenomenon of nucleolar dominance in plants, where nucleoli form around rRNA genes inherited from one parent only in one of the earliest described examples of epigenetic phenomena (Navashin, 1934). Chen and Pikaard (1997) found that growth in the presence of demethylating drug, 5-azaC resulted in a derepression of the previously silenced plant rRNA genes. This activation was also observed following treatment with the histone deacetylase inhibitors, trichostatin A and sodium butyrate. These results taken together suggest that DNA methylation and histone deacetylation are linked phenomena in transcriptional silencing.

### **1.8.1 DNA methylation recruits histone deacetylation**

Compelling support now exists that m<sup>5</sup>CpG-mediated silencing arises through recruitment of a repressor that directs the formation of a multiprotein repression complex containing chromatin-remodeling activities (Nan et al., 1998; Jones et al., 1998). Using deletions of GST-MeCP2 fusions, a region co-linear with the

transcriptional repression domain (TRD) associates with the co-repressor mSin3A and recruits histone deacetylases HDAC1 and HDAC2. Association between native MeCP2 and the mSin3 co-repressor complex was demonstrated by using anti-MeCP2 and anti-mSin3 antibodies to co-immunoprecipitate proteins from rat brain nuclear extracts. The proposed association between MeCP2 and mSin3A was further substantiated by immunoprecipitating histone deacetylase activity with anti-MeCP2 antibodies. The mSin3A interacting region of MeCP2 encompassing the TRD represses transcription in mouse L929 cells (Nan et al., 1997). This silencing was found to be partially derepressed following addition of TSA suggesting that at least part of the mechanism of gene inhibition was mediated by recruitment of the HDACs to deacetylate nucleosomal core histones (or other protein targets). Jones et al., (1998) tested directly whether the chromatin alterations and silencing observed on methylated hsp70 templates injected into *Xenopus* oocytes could be alleviated by treatment with TSA. DNaseI hypersensitivity was restored and the methylation-directed silencing could be abolished following inhibition of histone deacetylase activity. Antibodies to *Xenopus* MeCP2 and mSin3A showed these activities to exist in high molecular weight complexes that co-fractionated with each other further suggesting these proteins physically interacted. These findings by Nan et al., (1998) and Jones et al., (1998) establish a causal link between methylation-dependent gene inhibition, histone deacetylation and chromatin modification. Eden et al., (1998) have demonstrated that DNA methylation operates to induce histone hypoacetylation. Using cell lines containing stably integrated copies of methylated or unmethylated



HSV-tk, immunoprecipitation was carried out to distinguish between transcriptionally competent hyperacetylated mononucleosome fractions and bulk chromatin. This revealed that hyperacetylated regions preferentially contain the unmethylated and active construct whereas the methylated and inert gene was not found to be enriched in the highly acetylated nucleosomal fractions. These results suggested that the covalent modification by methylation of the reporter genes directed histone deacetylation because identical DNA sequences were used in the initial transfections. To assess whether the deacetylated histones preferentially associating with the methylated and silenced *tk* gene were responsible for the repression, the cells were treated with and without TSA. Treatment of the cells with TSA resulted in an increase in acetylation, increase in DNaseI sensitivity and importantly an elevation of transcription by the methylated tk construct. These three studies (Nan et al., 1998; Jones et al., Eden et al., 1998) strongly support the idea that DNA methylation directs the formation of an inactive chromatin structure by recruitment of histone deacetylase activities resulting in alteration of the acetylation status of the core histones H3 and H4 (Pazina and Kadonaga, 1997) or other non-histone proteins leading to transcriptional silencing.

### **1.8.2 Chromatin remodeling complexes and DNA methylation**

A growing number of distinct nucleosome remodeling complexes have been identified in eukaryotes (Travers, 1999). The RSC (remodeling the structure of chromatin) was identified in yeast on the basis of homology to the SWI/SNF complex (Cairns et al.,

1996). RSC possesses DNA-dependent ATPase activity and can remodel nucleosome structure in vitro however how the energy in ATP is used by RSC, SWI/SNF and the other nucleosome-remodeling activities is unclear at present. It has been suggested that an ATP-mediated translocation of these complexes on nucleosomal DNA would remodel local chromatin (Pazin et al., 1997). In *Drosophila*, the ATP-dependent nucleosome remodeling factor (NURF) is required for GAGA-mediated transcriptional activation (Tsukiyama et al., 1994; Tsukiyama and Wu, 1995). Interestingly, a subunit of the NURF complex, ISWI is homologous to the ATPase domain of the product of the *brahma* (*brm*) gene. The *brm* gene was identified as a dominant suppressor of mutations in *Polycomb* that exert positive and negative effects respectively on *Drosophila* homeotic gene expression during development. The ISWI subunit of NURF is also present in the CHRAC (chromatin accessibility complex; Varga-Weisz et al., 1997) and ACF (ATP-utilising chromatin assembly and remodeling factor; Ito et al., 1997) complexes which suggests that ATP hydrolysis by ISWI may lead to specific nucleosome remodeling within each distinct complex. ATP-dependent nucleosome remodeling was thought to be exclusively involved in transcriptional activation however the NuRD complex (nucleosome remodeling and histone deacetylase complex; Zhang et al., 1998) contains both ATP-dependent nucleosome remodeling activities and histone deacetylases. The purified NuRD complex has seven subunits, the SWI2/SNF2 helicase/ATPase domain-containing protein Mi2, the histone deacetylases HDAC1 and HDAC2, the histone-binding proteins RbAp46 and RbAp48, the metastasis-associated protein MTA2 and

the methyl CpG-binding domain family member, MBD3 (Zhang et al., 1999). The MBD3 protein is highly related to MBD2 which binds to methylated DNA in vitro and in vivo (Hendrich and Bird, 1998). However, mouse MBD3 does not bind methylated DNA with high affinity (Hendrich and Bird, 1998; Wade et al., 1999) and does not localise with the heavily methylated mouse major satellite when overexpressed (Hendrich and Bird, 1998). MBD2 was found to associate with the NuRD activity and target the complex to methylated DNA whereas the highly related MBD3 protein could not direct association of the NuRD complex to methyl-CpG specifically (Zhang et al., 1999). MBD2 is able to function as a transcriptional repressor when tethered to a cis-linked reporter gene and this silencing is derepressed completely in certain cases following addition of trichostatin A demonstrating that the mechanism of inhibition is deacetylase dependent (Ng et al., 1999). Dramatically, Anti-MBD2 and anti-HDAC1 antibodies were found to retard the electrophoretic mobility of the MeCP1 complex. Furthermore, anti-HDAC1 antibodies can immunodeplete the MeCP1 and MBD2 activities from HeLa cells strongly suggesting that MBD2 is a component of the MeCP1 methylation-specific repressor complex. MBD2 has been reported to demethylate DNA in vitro which would result in gene activation (Bhattacharya et al., 1999) and reverse the histone deacetylase-mediated gene silencing, however the MBD2 demethylase activity has not been independently verified (Zhang et al., 1999; Ng et al., 1999; Wade et al., 1999). These findings establish a second methylation-dependent, deacetylase-mediated transcriptional repression complex in mammalian cells in addition to the MeCP2-Sin3 histone

deacetylase complex. MeCP1 binds multiple symmetrically methylated sites (Meehan et al., 1989) whereas MeCP2 can bind a single methyl-CpG pair (Nan et al., 1993). These differences in methyl CpG-binding specificities between MeCP1 and MeCP2 may serve to mediate distinct transcriptional silencing requirements.

Given the absence of the Sin3 co-repressor that serves to nucleate HDAC1/2-containing complexes on DNA-binding proteins (Alland et al, 1997; Hassig et al., 1997; Heinz et al., 1997; Kadosh and Struhl, 1997; Nagy et al., 1997; Zhang et al., 1997) within the NuRD complex, the Mi2-HDAC1 interaction was analysed by pull down assays and found to require the PHD-zinc fingers of Mi2 but was not itself sufficient for association between Mi2 and the HDACs (Zhang et al., 1998) suggesting that PHD-zinc finger proteins interact indirectly with the histone deacetylase machinery.

### **1.8.3 Conservation of the NuRD complex during evolution**

The NuRD complex is evolutionarily conserved and represents the most abundant form of deacetylase activity in *Xenopus* oocytes (Wade et al., 1998) as well as in cultured mammalian cells (Zhang et al., 1998). Both *Drosophila* and *C. elegans* encode proteins homologous to the metastasis-associated protein 1, (Wade et al., 1999; Solari et al., 1999). Furthermore an MBD2/3 homolog has been identified in *Drosophila* (Wade et al., 1999) however no MBD2/3-related sequences have been characterised in the *C. elegans* genome to date. *Drosophila* Mi2 (dMi2) was

identified in a screen for proteins that interact with Hunchback (Hb) (Kehle et al., 1998). Gap proteins, like Hb are required to delimit the expression boundaries of the homeotic genes however they are transient and require Polycomb-group gene products to maintain the epigenetic gene silencing. This interaction between dMi2 and Hb suggests importantly that the silencing of homeotic genes by the Polycomb group requires histone deacetylation and furthermore that proteins with restricted DNA-binding specificities, compared to MBD2, are also capable of targeting the NuRD-deacetylase complex.

The *C. elegans egl-27* and *egr-1* genes are functionally redundant to one another and are homologous to MTA1 (Solari et al., 1999). Inhibition of *egl-27* resulted in a defect in patterning in certain regions during development whereas *egr-1* inhibition has a weak posterior defect. Strikingly, the double mutant contains global defects in tissue organisation however differentiation is unaffected. Interestingly, the sequence of EGL-27 shares a domain (ELM1: EGL-27 and MTA1 homology) present in the Dnmt1 maintenance methyltransferase and the origin recognition complex (ORC-1) protein which has lead to the suggestion that this motif might be involved in linking DNA methylation and replication (Callebut et al., 1999).

#### **1.8.4 Reversal of methylation-mediated silencing**

At a methylated, hypoacetylated and silenced gene locus there are a number of possibilities as to how this inhibition might be reversed : active demethylation would

result in MeCP displacement and attenuation of histone deacetylase activity, allowing histone acetyltransferase access to modify histones H3 and H4 to render the chromatin template accessible to transcription factors. Alternatively, as methylation by itself does not decrease transcription, it is conceivable that the histone acetyltransferases are capable of over-riding the methylation-directed deacetylase-associated repression. In support of this hypothesis, the tat transactivator is able to circumvent the methylation-mediated silencing at HIV promoter sequences (Bednarik et al., 1990) and TSA-mediated inhibition of histone deacetylase activity can allow transcriptional activation with methylation in place (Sheridan et al., 1997). This model where methyl-CpG binding proteins recruit co-repressors to silence gene activity at methylated promoters by alterations in chromatin structure obviously does not exclude the existence of additional methyl-CpG binding proteins that serve to mediate similar effects. The possibility that additional mechanisms whereby CpG methylation could reduce transcription also remains open. In mouse L cells and *Xenopus* oocytes, TSA did not lead to a complete restoration of transcriptional competence of the reporter repressed by MeCP2 and as MeCP2 is undetectable in HeLa cells (Ng and Bird, 1999) which are known to efficiently repress methylated templates this suggests the existence of additional mediators of the global effects of DNA methylation on chromatin and gene expression.

## **1.9 MeCPs and expression inheritance**

In the mammalian genome the majority of CpG sites, occurring every 100bp or so, in bulk intergenic DNA are methylated whereas the remaining fraction reside within CpG islands that are constitutively unmethylated (Bird, 1986). CpG islands in housekeeping genes remain free from methylation, whereas genes exhibiting a tissue-restricted pattern of expression are generally methylated in every cell with the exception of tissues that express the gene. Pollack et al., (1980) first demonstrated that this methylation pattern was clonally inherited and the Dnmt1 maintenance methyltransferase is the proposed enzymatic activity that maintains the methyl-CpG content on nascent hemimethylated daughter strands following DNA replication (Holliday, 1987). Dnmt1 possesses an N-terminal domain that is involved in targeting to replication foci and the re-establishment of the symmetric methylation at CpG sites occurs rapidly following de novo DNA synthesis (Leonhardt et al., 1992; Gruenbaum et al., 1983) whereas the assembly of chromatin takes place later (Cusick et al., 1983). Little is known how cell-specific expression patterns based on nucleosome positioning and chromatin structure occurs and how they are inherited following chromatin disruption at the replication fork after DNA synthesis. However an attractive possibility is that the stably, inherited methylation pattern directs the formation of inert chromatin structures that are transcriptionally opaque through binding of MeCP1 and MeCP2 plus additional methylation-specific silencers that recruit histone deacetylase co-repressors. Evidence suggests that the default state of nascent chromatin is hyperacetylated as newly synthesised histones are acetylated

and deposited onto newly replicated DNA (Sobel et al., 1995). The potent activator GAL4-VP16 can transcribe methylated templates if bound prior to assembly into chromatin (Kass et al., 1997). Macleod et al., (1994) and Brandeis et al., (1994) provided support for a model where binding sites for the transactivator Sp1 and presumably active gene transcription generally (Macleod et al., 1998) is necessary to maintain an unmethylated CpG island. DNA demethylation may then conceivably occur by a passive mechanism through the steric exclusion of the maintenance methyltransferase, Dnmt1 at CpG islands that are transcriptionally active (Matsuo et al., 1998). Thus the propagation of DNA methylation on newly synthesised DNA could help establish the acetylation profile to control gene accessibility where active genes would remain methylation free and contain hyperacetylated histones H3 and H4 (Tazi and Bird, 1990) whereas methylated sequences would direct via methyl-CpG binding proteins a hypoacetylated state (Nan et al., 1998; Jones et al., 1998; Eden et al., 1998) present throughout the cell cycle. The resistance of methylated domains to nuclease treatment is known to increase appreciably following DNA replication, consistent with this idea and high-densities of DNA methylation can prevent V(D)J recombination in lymphocytes after replication has occurred (Hsieh and Lieber, 1992). The ability to maintain epigenetic expression states may well be dependent on this ability of DNA methylation to pattern histone acetylation in the genomes of higher vertebrates.



## **CHAPTER TWO: MATERIALS AND METHODS**

## **2.1 Bacterial Cell Culture**

### **2.1.1 Media and solutions**

All media were sterilised by autoclaving or by filtration through a 0.22µm membrane before use.

#### L-broth (LB) agar and top agarose

10g Bacto-tryptone (Difco), 5g yeast extract (Difco), 10g NaCl and 2.46g MgSO<sub>4</sub> were added per litre of dH<sub>2</sub>O. L-broth agar contained in addition 15g/L agar (Oxoid) and top agarose 7g/L agarose (Boehringer Mannheim).

#### 2X YT broth

10g yeast extract, 5g NaCl and 16g Na<sub>2</sub>HPO<sub>4</sub> per litre dH<sub>2</sub>O.

#### SOC broth

20g tryptone, 5g yeast extract, 0.59g NaCl, 1.86g KCl, 10mM MgCl<sub>2</sub>, 10mM MgSO<sub>4</sub> and 20mM glucose per litre dH<sub>2</sub>O.

#### Media additives

Ampicillin was used at 50µg/ml and tetracycline at 12.5µg/ml where selection was appropriate. X-gal and IPTG were added to molten L-agar to facilitate identification of recombinant bacteria by α-complementation. To prepare plating cells for adsorption of λ phage, cells were grown in 0.2% maltose and 10mM MgSO<sub>4</sub> to increase competency. Bacteria were grown typically at 37°C with good aeration. Bacterial stocks were maintained at -80°C in LB media containing 20% glycerol.

### 2.1.2 Bacterial strains

Bacterial hosts for transformation and propagation of plasmid episomes were either *E. coli* strain :

XL1 Blue - supE44, hsdR17, recA1, endA1, gyrA96, thi, relA1lac<sup>-</sup>, F' : pro AB<sup>+</sup>, lacI<sup>q</sup>, lacZM15, Tn10(tet<sup>r</sup>) or

DH5 $\alpha$  - supE44, hsdR17, recA1, endA1, ghrA46, this-1 relA1

For protein expression the bacterial host was *E. coli* strain :

BL21(DE3)pLysS - F' : dcm ompT hsdS(r<sub>b</sub><sup>-</sup> m<sub>b</sub><sup>-</sup>) gal  $\lambda$  (DE3) [pLysS Cam<sup>r</sup> ]

### 2.1.3 Preparation of competent cells for electroporation

1 litre of L broth was inoculated with 1ml of a fresh overnight culture (either XL1-Blue or DH5 $\alpha$ ) and grown with vigorous shaking at 37°C to an OD<sub>595nm</sub> of approximately 0.7 - 0.8. The cells were harvested by chilling on ice and pelleting at 3000g for 10 minutes at 4°C. The cells were resuspended in 500mls of ice cold dH<sub>2</sub>O (or other low ionic strength buffer) and repelleted by centrifugation. The cells were then subsequently washed twice in 100ml of ice cold dH<sub>2</sub>O, 10ml of 10% glycerol and repelleted by centrifugation. The cells were then resuspended in a final volume of 2-3ml of 10% glycerol, aliquoted as 40 $\mu$ l stocks and stored at -80°C at a cell concentration of approximately 3 X 10<sup>10</sup> cells/ml.

## **2.1.4 Electroporation of competent cells**

An aliquot of competent cells were thawed on ice and plasmid DNA or ligation reaction were added, mixed and left on ice for 1 minute. This mixture was then transferred to a chilled, sterile electroporation cuvette and subjected to a pulse of 2.5kV in a Gene Pulser (BioRad). The time constant for the pulse should be between 4.5 and 5.0 msec. 1ml of SOC was immediately added to the cells and resuscitated at 37°C with constant shaking at 250 RPM for 1hr before plating on L-agar plates containing antibiotic selection where appropriate and incubated overnight at 37°C.

## **2.2 Cell and Tissue Culture**

### **2.2.1 Media and reagents**

#### DMEM (Gibco BRL)

Dulbecco's Modified Eagle Medium supplemented with 10% Bovine calf serum (GIBCO BRL) and penicillin/streptomycin at 100 units/ml and 100g/ml respectively.

#### MEM Alpha (Gibco BRL)

MEM Alpha medium supplemented with 10% Bovine calf serum (GIBCO BRL) and penicillin/streptomycin at 100 units/ml and 100g/ml respectively.

#### Trypsin/EDTA solution (T/E)

2.5% trypsin and 0.02% EDTA in phosphate buffered saline (PBS).

### **2.2.2 Mammalian cell lines**

EFS2 cells were derived from an murine ES cell line containing a randomly integrated  $\beta$ -geo transgene (Hendrich and Bird, 1998).

HeLa cells are derived from a cervical carcinoma expressing the human papiloma virus 16 E6 tumour antigen.

L929 (L) cells are a mouse fibroblast cell line.

### **2.2.3 Thawing of cells stored in liquid nitrogen**

All cell lines utilised in this work were stored in liquid nitrogen, in 1ml of culture medium and 10% DMSO. Cells were recovered by rapidly thawing in a 37°C water bath, then resuspending in 9mls of the appropriate medium and spinning at 1200 RPM for 2 minutes. The supernatant was then aspirated and the pellet resuspended gently in 10mls of fresh medium in a T-25 flask at 37°C under 5% CO<sub>2</sub> atmosphere. The culture medium was changed 24 hrs after plating out the cells to wash out residual DMSO.

### **2.2.4 Maintenance of cell lines**

Cell culture work was carried out in a laminar flow hood.

L929 cells and HeLa lines were cultured in DMEM, supplemented with 10% BCS and penicillin/streptomycin at 100 units/ml and 100g/ml respectively. These lines were grown as a monolayer in sterile T-75 flasks in a 37°C incubator under 5% CO<sub>2</sub> atmosphere. These lines were passaged once a week by washing the cells twice in

PBS, adding 2mls of T/E for 2 min to dislodge the cells, pelleting gently at 1200RPM X 3 min and then re-seeding at 1:10 dilution to maintain the line.

EFS2 fibroblasts were cultured as a monolayer in MEM alpha medium supplemented with 10% BCS and penicillin/streptomycin at 100units/ml and 100g/ml respectively and this line was passaged twice weekly.

### **2.2.5 Transfection of mammalian cell lines**

In order to establish transient cell lines, cells were grown in either 3mm dishes or six well culture dishes until 30-50% confluent then transfected with constructs by either of the following methods.

DNA-liposome transfection: 10 $\mu$ l of either Lipfectin or Lipofectamine (Life Technologies) were diluted in 100 $\mu$ l of MEM Alpha medium without serum and allowed to incubate at room temperature for 45 min. Various concentrations of DNA (500ng - 10 $\mu$ g) were resuspended in 100 $\mu$ l of MEM Alpha medium and the two solutions were gently mixed and complex formation allowed to proceed for 15 min. Subsequently 0.8ml of serum-free and antibiotic-free MEM Alpha medium was added to the DNA-lipofectin mix and overlaid onto cells that had been washed with serum-free medium. Cells were incubated for 5 hr at 37<sup>0</sup>C incubator under 5% CO<sub>2</sub> atmosphere and the DNA containing medium was removed and replaced with complete growth medium and incubated for a further 48 hr. Cells were then either

fixed in 4% paraformaldehyde in phosphate-buffered saline for 20 min at 37°C and then washed twice in phosphate-buffered saline prior to counterstaining or assayed as in 2.2.6.

DEAE-dextran transfection: Approximately  $5 \times 10^5$  mouse L cells (1:10 split) were grown in culture until 30-50% confluent. Prior to transfection, DNA (5-10 $\mu$ l) in TE was added to 30 $\mu$ l of TBS and subsequently added to 80 $\mu$ l of 10mg/ml DEAE-dextran in TBS while constantly shaking the tube. The media was aspirated from the L929 cells and the plate washed with PBS and overlaid with 1ml of DMEM-1%FCS. The DNA/DEAE-dextran mixture was then added dropwise to each plate equally over the surface of the cells and incubated for 4 hr in a 37°C incubator under 5% CO<sub>2</sub> atmosphere. This DNA/DEAE-dextran mixture was then aspirated from the plates and the cells were shocked by adding 5ml of 10% DMSO in PBS for 1 min. The DMSO was aspirated, washed with PBS twice and 1ml of complete medium was added to each plate. Cells were incubated for 48 hr and then assayed as in 2.2.6.

### **2.2.6 $\beta$ -galactosidase assay**

Cells were grown to confluence following transfection of reporter constructs, washed twice in phosphate-buffered saline and overlaid with 500 $\mu$ l of 1X lysis buffer (Promega) and incubated for 15 min at room temperature. Cells were carefully scraped down using rubber policemen and the suspension was transferred to a sterile microfuge tube and put on ice. Samples were subsequently centrifuged at 14000

RPM for 2 min at 4°C and 150µl of the supernatant was assayed directly by adding an equal volume of 2X Assay Buffer (120mM Na<sub>2</sub>HPO<sub>4</sub>, 80mM NaH<sub>2</sub>PO<sub>4</sub>, 2mM MgCl<sub>2</sub>, 100mM β-mercaptoethanol, 1.3 mg/ml ONPG) and incubated at 37°C for 30 min. The reaction was then terminated by addition of 1M sodium carbonate and the spectrophotometric absorbance at 420 nm determined.

## **2.3 Preparation of Nucleic Acids**

### **2.3.1 Isolation of plasmid DNA**

10mls of LB media containing 50µg/ml ampicillin (Sigma) were inoculated with a single bacterial transformant and incubated with shaking overnight at 37°C. Cells were harvested by centrifugation at 2000 RPM for 10 min. Cells were resuspended in 250µl 50mM Tris-HCl, 10mM EDTA, RNase A 10µg/ml, then lysed in 250µl 0.2M NaOH, 1% SDS and the suspension was mixed thoroughly by inversion before adding 300µl of a neutralising solution containing 3M potassium acetate, 2M acetic acid. Bacterial debris was then pelleted at 14000 RPM for 10 min and the supernatant was applied to a column containing a selectively absorptive silica based resin (Qiagen). Plasmid DNA was bound by centrifugation for 14000 RPM for 1 min and washed in 750µl 55% ethanol, 8.0mM Tris-HCl (pH 8.0) and 40mM EDTA, then spun twice in a microcentrifuge at 14000 RPM for 1 min to remove residual ethanol. The bound plasmid DNA was eluted in either dH<sub>2</sub>O or TE (10 mM Tris-HCl, pH 7.5, 1mM EDTA).



### **2.3.2 Isolation of total RNA from cultured cells and tissues**

This protocol is based on the single step method for total RNA isolation (Chomczynski and Sacchi, 1987). Typically 100mg of tissue (frozen in liquid nitrogen after dissection) or cells harvested from a near confluent 25cm<sup>3</sup> tissue culture flask were homogenised in 500µl of Solution D (4M guanidinium isothiocyanate, 25mM sodium citrate pH 7.0, 0.5% L-lauryl sarcosine, 0.1M β-mercaptoethanol). 20µl of DEPC-treated 2M sodium acetate, 500µl of water-saturated phenol and 100µl of chloroform were added and the mixture was vigorously shaken and placed on ice for 20 min. After centrifugation at 14000 RPM for 20 min at 4<sup>0</sup>C the RNA-containing supernatant was precipitated with one volume of isopropanol at -20<sup>0</sup>C for at least one hour. RNA was pelleted by centrifugation again at 14000 RPM for 20 min at 4<sup>0</sup>C and the pellet was resuspended in 300µl of solution D and reprecipitated with one volume of isopropanol at -20<sup>0</sup>C. The precipitated RNA was washed in 75% ethanol, repelleted, dried and either resuspended in DEPC-treated dH<sub>2</sub>O or stored as ethanol precipitates at -70<sup>0</sup>C. Yield and relative purity were assessed by measuring the absorbance of a diluted aliquot of this total RNA at wavelengths of 260nm and 280nm.

### **2.3.3 Synthesis of cDNA for RT-PCR**

5µg of total RNA was adjusted to a 12µl volume containing 250ng of random hexanucleotide primers (Boehringer Mannheim) with DEPC-treated dH<sub>2</sub>O. This sample was heated to 70<sup>0</sup>C for 10 min and cooled on ice. To this solution was added

2 $\mu$ l of 10X PCR buffer, 2 $\mu$ l of MgCl<sub>2</sub>, 1 $\mu$ l of 10 mM dNTPs and 2 $\mu$ l 0.1M DTT per reaction. The sampled was mixed gently and incubated at 25<sup>0</sup>C for 10 min followed by incubation at 42<sup>0</sup>C for 2 min. 1 $\mu$ l (200 units) of Superscript II reverse transcriptase (GIBCO BRL) was added to the reaction and incubated at 42<sup>0</sup>C for a minimum of 50 min. The reaction was stopped by heat inactivation at 70 for 15 min. The cDNA was now used directly in RT-PCR experiments (2 $\mu$ l per reaction).

## **2.4 Gel Electrophoresis**

### **2.4.1 Electrophoresis of DNA**

For both preparative and analytical purposes 0.8% to 1.5% (w/v) agarose gels in 1X TAE (40mM Tris-acetate, 1mM EDTA) containing 1 $\mu$ l of ethidium bromide (10mg/ml) per 50mls melted agarose were used to separate DNA fragments ranging in size from 300bp to 15kb. Samples were prepared by addition of 0.1V of loading buffer (15% Ficoll type 400, 10mM EDTA and 0.25% orange G) and following electrophoresis in 1X TAE buffer, DNA was visualised on a UV transilluminator.

### **2.4.2 Electrophoresis of RNA**

For Northern blots or in order to assess total RNA quality prior to cDNA synthesis, RNA samples were run on horizontal 1.2% agarose gels in 1X MOPS (1M MOPS pH 7.0, 3M sodium acetate pH 7.0, 0.5M EDTA) containing 1.1% formaldehyde. To 10 $\mu$ g of total RNA in 6 $\mu$ l of DEPC-treated dH<sub>2</sub>O, 12.5 $\mu$ l of deionised formamide, 2.5 $\mu$ l of 10X MOPS, 4 $\mu$ l of formaldehyde and 1 $\mu$ l ethidium bromide. Samples were heated to 65<sup>0</sup>C for 10 min, 0.1V of RNA loading buffer (15% Ficoll, and 0.25%

bromophenol blue and 0.25% xylene cyanol FF) added, and following electrophoresis in 1X MOPS buffer, RNA was visualised on a UV transilluminator.

## **2.5 Radioisotopic labelling of DNA fragments**

### **2.5.1 Random Primed Labelling of DNA probes**

DNA probes for blot hybridisation were labelled with [ $\alpha$ - $^{32}$ P]-dCTP by random hexanucleotide primers using the Klenow fragment of *E. coli* DNA polymerase I (Feinberg and Vogelstein, 1983). 50ng of DNA probe was resuspended in 12 $\mu$ l dH<sub>2</sub>O, denatured by boiling for 5 min and chilled on ice. To this 4 $\mu$ l of 5X oligolabelling buffer (0.25M Tris-HCl, 50mM MgCl<sub>2</sub>, 0.5mM dithioerythritol, 1mg/ml BSA, 312 A260 units/ml hexanucleotides, 250 $\mu$ M each of dATP, dTTP, dGTP pH 7.2), 30 $\mu$ Ci [ $\alpha$ - $^{32}$ P]-dCTP (10 $\mu$ Ci/ $\mu$ l) (Amersham) and finally 1 $\mu$ l Klenow enzyme (2 units) was added and incubated at 37°C for 30 min. Unincorporated nucleotides were removed by centrifugation through a 1ml Sephadex G-50 column (Pharmacia). Prior to hybridisation the probe was denatured by boiling for 5 min, cooled immediately on ice and added to the hybridisation mix.

### **2.5.2 End Labelling of DNA oligonucleotides**

Oligonucleotides were labelled by the transfer of the [ $\gamma$ - $^{32}$ P]-ATP to the terminal 5'-OH group of the oligonucleotide using polynucleotide kinase (PNK). To 50ng of oligonucleotide in 7 $\mu$ l of dH<sub>2</sub>O, 2 $\mu$ l of PNK buffer (Boehringer Mannheim), 10 $\mu$ l of (10 $\mu$ Ci/ $\mu$ l) [ $\gamma$ - $^{32}$ P]-ATP. The reaction was incubated at 37°C for 40 min. It was added

directly to the hybridisation mix without the need for further removal of incorporated nucleotides.

## **2.6 Nucleic Acid Blotting and Hybridisation**

### **2.6.1 Southern blot analysis of DNA**

Southern blotting was used to transfer DNA from agarose gels to nylon membranes by capillary action (Southern, 1975). Following electrophoresis, agarose gels were soaked in 0.4M NaOH for 20 min and transferred to 3MM filter paper using 0.4M NaOH as blotting buffer. The gel was overlaid with a Hybond N+ nylon membrane (Amersham) and transfer of DNA was allowed to proceed overnight. Following transfer the filter was washed briefly in 5X SSC (20X SSC : 175.3g NaCl, 88.2g sodium citrate pH7.0 per litre) and pre-hybridised in Church-Gilbert buffer (0.5M sodium phosphate pH 7.2, 7% SDS, 1mM EDTA, 0.5% powdered skimmed milk) for at least one hour prior to hybridisation (Church and Gilbert, 1984). Random primed double stranded probes were denatured by boiling for 5 min, added to 25ml of fresh Church Gilbert buffer and allowed to hybridise overnight. Membranes were washed following hybridisation three times with 100ml of 0.2X SSC, 0.1% SDS for 30 min each. Membranes were sealed in plastic wrap and exposed to X-ray film (XAR5 Kodak). Hybridisations and subsequent washes were carried out at 65°C for random primed probes or at 5°C below the  $T_m$  of end labelled oligonucleotide probes.

### **2.6.2 Northern Blot Analysis of RNA**

10µg of total RNA was electrophoresed on denaturing formaldehyde gels and washed in dH<sub>2</sub>O to remove formaldehyde. RNA was transferred to nylon membranes as described for Southern blotting (see 2.6.1) except that the transfer buffer was 20X SSC. Following overnight transfer, the RNA was immobilised by baking the filter at 80°C and crosslinked with UV radiation (312nm). Membranes were hybridised, washed and exposed exactly as described for DNA hybridisation.

### **2.6.3 Bacteriophage plaque lifts**

Bacteriophage plaque lifts were carried out as duplicates using Hybond N<sup>+</sup> nylon membranes (Amersham). Nylon membrane filters were placed on pre-chilled top agarose plates for 1 min and marked by puncturing the membrane asymmetrically to facilitate subsequent orientation. Recombinant λ bacteriophage adsorbed on membranes were denatured by placing plaque side up on 3MM filters soaked in 1.5M NaCl, 0.5M NaOH for seven min, then neutralised by placing membranes on 3MM filters soaked in 1.5M NaCl, 0.5M Tris-HCl pH 7.5 for 2 min and this neutralisation step was repeated. Filters were then washed in 2X SSC to remove any top agarose, air dried and baked at 80°C for 2 hours prior to hybridisation of radiolabelled probes.

## **2.6.4 Removal of Probe Following Hybridisation**

In order to hybridise the same membranes with alternative probes, hybridised probe was removed by pouring boiling 0.5% SDS over the membrane and allowing to cool to room temperature with shaking. This was performed twice and the membrane was exposed to X-ray film to confirm probe removal.

## **2.7 Enzymatic Manipulation of DNA**

### **2.7.1 Restriction endonuclease digestion of DNA**

Restriction endonuclease digestion was carried out in the appropriate buffer at the temperature recommended by the manufacturer (New England Biolabs). Typically, a 2-3 fold unit excess of enzyme was employed to restrict DNA substrates.

### **2.7.2 Dephosphorylation of DNA**

In some cases to prevent self-ligation of digested vectors, the 5' phosphate group was removed by dephosphorylation with calf intestinal phosphatase (CIP). Following cleavage of vector DNA and gel isolation, DNA was dephosphorylated by CIP (GIBCO BRL) in appropriate buffers as recommended by the manufacturer.

### **2.7.3 Ligation of DNA**

DNA ligase catalyses the ligation of DNA ends by the formation of a phosphodiester bond between the 3' hydroxyl group of one end and the 5' phosphate of another. Ligation reactions were normally carried out in a 20 $\mu$ l reaction volume with a molar ratio of insert to vector of 3:1 in the presence of 50mM Tris-HCl, 10mM MgCl<sub>2</sub>,

10mM DTT, 1mM ATP, 50µg/BSA and 100 units of T4 DNA ligase (NEB) at 16<sup>0</sup>C overnight. Ligation reactions were spotted onto filters (Millipore type VS, 0.025µm) and desalted by microdialysis against dH<sub>2</sub>O for 30 min prior to electrotransformation.

## **2.8 Polymerase Chain Reaction**

The polymerase chain reaction (PCR) (Saiki et al., 1988) was employed in order to amplify DNA fragments for a variety of applications. Reactions were carried out in a 50µl volume containing 10mM Tris-HCl pH 8.2, 50mM KCl, 0.2mM dNTPs, 1µM of each primer with 100ng of template and 2 units of Taq polymerase (Qiagen) in a PCR cycler (Hybaid) with heated lids. The polymerase used was dependent on the nature of the downstream application so where fidelity was the primary criterion proof-reading enzymes, such as *Pfu* (Stratagene) containing 3'→5' exonuclease activity were employed and where yield was most important reactions were cycled with conventional non-proof reading polymerases. DNA was amplified by initial denaturation for 2 min at 94<sup>0</sup>C, then 30 cycles of 94<sup>0</sup>C for 15 sec, 55<sup>0</sup>C for 30 sec, then 72<sup>0</sup>C for 1 min per kb of product size and a final extension of 72<sup>0</sup>C for 10 min.

## **2.9 DNA Sequencing**

DNA sequencing was carried out by sequence cycling using fluorescent dye terminators (Perkin-Elmer). Sequencing reactions were carried out in 10µl volume containing 0.5µg of template DNA, 3.2pmol of primer and 4µl of terminator pre-mix

(A-dye terminator, T-dye terminator, C-dye terminator, G-dye terminator, dATP, dTTP, dCTP, dITP, Tris-HCl pH 9.0, MgCl<sub>2</sub>, thermostable pyrophosphatase and AmpliTaq DNA polymerase). Reactions were sequence cycled in a PCR machine (Normal : 96<sup>0</sup>C for 30 sec, 50<sup>0</sup>C for 20 sec, 60<sup>0</sup>C for 4 min for 25 cycles; GC-rich : 98<sup>0</sup>C for 30 sec, 55<sup>0</sup>C for 20 sec, 60<sup>0</sup>C for 4 min for 25 cycles). Samples were precipitated by addition of 1µl of 3M sodium acetate and 25µl of 95% ethanol on ice in the dark for 10 min. Reactions were then centrifuged at 14000 RPM followed by washing of the pellet in 75% ethanol and drying. Samples were dissolved in 4µl of loading buffer (25mM EDTA/50mg/ml blue dextran and a 1:5 ratio of de-ionised formamide), heated to 90<sup>0</sup>C for 2 min and loaded on polyacrylamide gels in a 373 automated DNA sequencer (Applied Biosystems) operated by Vicky Clark. Sequence files were analysed using Seqman (DNA star) and GeneJockey (Biosoft) software.

## **2.10 Screening of cDNA libraries**

### **2.10.1 Preparation of phage-competent cells and library titration**

Single colonies of XL1-Blue were inoculated into 10 ml of L-broth containing 0.2% maltose and 10mM MgSO<sub>4</sub> and grown overnight at 37<sup>0</sup>C to generate bacteriophage-competent host bacterial cells. cDNA libraries were titred by serial dilution and for large screens 3 X 10<sup>5</sup> pfu of bacteriophage were incubated with 0.8ml of bacteriophage-competent bacteria for 20 min and were plated onto pre-warmed 22.5 X 22.5 cm plates in 40ml top agarose and incubated overnight at 37<sup>0</sup>C. Bacteriophage plaque lifts were carried out as in 2.6.3. Secondary screening were carried out on all



positives at a density of approximately 100 pfu per 175 cm<sup>2</sup> plate using 0.6ml of XL1 blue host bacterial cells and 10ml of top agarose. Following tertiary screening individual strongly hybridising plaques were picked and transferred to a sterile microfuge tube containing 500µl SM buffer (Per litre : 5.8g NaCl, 2g MgSO<sub>4</sub>.7 H<sub>2</sub>O, 50ml 1M Tris.Cl pH 7.5, 5 ml 2% gelatin solution) and 20µl chloroform. The cored plaques of interest were vortexed briefly to release the λZAP II phage particles into the SM buffer and incubated at room temperature for 1-2 hrs or overnight at 4<sup>0</sup>C. This phage stock is stable for one year at 4<sup>0</sup>C (Stratagene).

#### **2.10.2 In vivo excision of cDNA inserts from λZAP II cDNA libraries**

In a 50 ml conical tube 200µl of OD<sub>600</sub> = 1.0 XL1 blue cells, 200µl of phage stock (>1 X 10<sup>5</sup> pfu/ml) and 1µl of R408 helper phage (>1 X 10<sup>6</sup> pfu/ml) were incubated at 37<sup>0</sup>C for 15 min. 5 ml of 2X YT media was added and incubated at 37<sup>0</sup>C for 3 hr. The tube was heated at 70<sup>0</sup>C for 20 min, centrifuged for 5 min at 4000 g and then 10µl of the supernatant was added to 200µl of OD<sub>600</sub> = 1.0 XL1 blue cells and incubated at 37<sup>0</sup>C for 15 min. 1µl to 100µl was spread on LB plates containing ampicillin and incubated overnight at 37<sup>0</sup>C. Plasmid DNA from transformants was prepared as in 2.3.1, restriction mapped and inserts sequenced using the T3 and T7 promoters in pBluescript SK.

## **2.11 Fluorescent *In Situ* Hybridisation (FISH)**

### **2.11.1 Labelling genomic probes with biotin-16-dUTP**

To incorporate biotin-16-dUTP to genomic probes a nick translation reaction was set up containing 2 $\mu$ l of 10X nick translation buffer (0.5M Tris pH 7.5, 0.1M MgSO<sub>4</sub>, 1mM DTT and 500 $\mu$ g/ml BSA Fraction V), 2.5 $\mu$ l of each of 0.5mM dATP, dCTP, dGTP and biotin-16-dUTP, 0.5 $\mu$ g of genomic probe, 1 $\mu$ l of DNaseI (Boehringer Mannheim) freshly diluted 1:500 and 10 units of DNA polymerase I (GIBCO BRL) adjusted to a final 20 $\mu$ l volume with dH<sub>2</sub>O. The reaction was incubated at 16<sup>0</sup>C for 90 min and then precipitated with 0.1 volume of 3M sodium acetate and 2 volumes of ethanol and resuspended in TE.

### **2.11.2 Detection of biotin-16-dUTP incorporation**

Botin-16-dUTP labelled probes were serially diluted and spotted onto Hybond N<sup>+</sup> filters (Amersham) with biotin labelled standards (1-20pg biotin; GIBCO BRL) and UV crosslinked at 250mJ. The filter was washed in Buffer I (0.1M Tris pH 7.5, 0.15M NaCl) for 5 min at room temperature and subsequently incubated in Buffer II (0.1M Tris pH 7.5, 0.15M NaCl, 3% BSA Fraction V) at 60<sup>0</sup>C for 1 hr. The filter was incubated in 10ml of Buffer I containing 10 $\mu$ l of a 1:1000 dilution of streptavidin alkaline phosphatase (Boehringer Mannheim) in 0.5M Tris pH 7.5, 0.15M NaCl for 10 min at room temperature. The filter was washed twice with Buffer I for 15 min and once in Buffer III for 5 min and finally incubated with 10 ml of Buffer III with

1ml of BCIP/NBT alkaline phosphatase substrate kit IV (Vector) for 3 hr and biotin-16-dUTP incorporation was estimated.

### **2.11.3 *In situ* hybridisation with biotin-labelled probes**

Approximately 100ng of biotin-labelled probe was precipitated with two volumes of ethanol and the pellet was resuspended in 10 $\mu$ l of hybridisation mix (50% formamide, 2X SSC, 1% Tween20, 10% dextran sulphate, 5 $\mu$ g salmon sperm DNA, 2 $\mu$ g human CotI DNA) at 37 $^{\circ}$ C for 3-4 hr. Slides with human metaphase spreads were incubated with 100 $\mu$ l of RNase A (100  $\mu$ g/ml 2X SSC) per slide for 1 hr at 37 $^{\circ}$ C. Slides were subsequently dehydrated in an ethanol series (70%, 90%, 100%) for 2 min each and allowed to dry. Slides were denatured by incubating in 70% formamide, 2X SSC for 3 min at 70 $^{\circ}$ C and then subsequently dehydrated through the ethanol series as before. Biotin-labelled probes were denatured at 70 $^{\circ}$ C for 5 min and then re-annealed for 15 min at 37 $^{\circ}$ C. 10 $\mu$ l of probe were added to each slide, covered in parafilm and incubated at 37 $^{\circ}$ C overnight. Slides were then washed for 4 X 3 min in each of the following solutions : 50% formamide / 2X SSC at 45 $^{\circ}$ C; 2X SSC at 45 $^{\circ}$ C; 0.1X SSC at 60 $^{\circ}$ C to remove non-specifically bound probe. 50 $\mu$ l of blocking buffer (4X SSC, 5% skimmed milk) was then applied to each slide and left at room temperature for 5 min. 50 $\mu$ l of avidin-FITC (Vector) was applied to each slide, covered in parafilm and incubated at 37 $^{\circ}$ C for 30 min. Slides were then washed three times in 4X SSC, 0.1% Tween 20 for 2 min at 45 $^{\circ}$ C. 50 $\mu$ l of biotinylated anti-avidin (Vector) was applied per slide and incubated at 37 $^{\circ}$ C for 30 min and washed as above. Finally 50 $\mu$ l of

avidin-FITC was applied and incubated and treated as before and allowed to dry. Slides were counter-stained with 20 $\mu$ l of DAPI (10 $\mu$ g/ml; Sigma) on coverslips and mounted in Vectashield (Vector) and images captured under an epifluorescence microscope (Zeiss).

## **2.12 Expression and affinity purification of recombinant proteins**

### **2.12.1 Affinity purification of histidine tagged recombinant proteins**

Overnight LB cultures of transformed *E. coli* BL21 DE3 (pLysS) were diluted 1:100 in 1 litre of fresh LB medium containing 50 $\mu$ g/ml ampicillin and grown at 37 $^{\circ}$ C until the OD<sub>600</sub> reached 0.5. The cultures were induced by addition of 0.4mM IPTG and incubated at 37 $^{\circ}$ C for a further 3-4 hrs. The induced cells were pelleted and washed twice in cold 50mM HEPES pH 7.9, 0.1M NaCl. Bacterial pellets were resuspended in 20 mls of resuspension buffer (20mM HEPES pH 7.9, 10% glycerol, 0.3M NaCl, 0.5mM PMSF, 10mM  $\beta$ -mercaptoethanol and the protease inhibitors leupeptin, antipain, chymostatin, pepstatin A, aprotin were included at 5 $\mu$ g/ml to minimise proteolysis). The cells were then lysed on ice by mild sonication and addition of Triton-X-100 to a final concentration of 0.1%. Samples were centrifuged at 16000 RPM at 0 $^{\circ}$ C for 30 min and 20mM imidazole was added to minimise non-specific binding during the subsequent purification steps. A Ni<sup>2+</sup>-NTA superflow column (Qiagen) was packed to 1.5ml and equilibrated with 10mls of resuspension buffer containing 0.1% Triton-X-100 and 20mM imidazole. The crude lysate was passed twice over the column, then washed with 20mls of resuspension buffer. Finally, the

bound proteins were eluted from the column in 6.0mls of resuspension buffer containing 0.4M imidazole, analysed by SDS-PAGE and stored at -80°C.

### **2.12.2 Affinity purification of glutathione-S-transferase (GST)-tagged recombinant proteins**

For expression of GST-fusion proteins, an overnight culture of transformed *E. coli* BL21 DE3 (pLysS) was diluted 1:100 in 1 litre of fresh LB medium containing 50µg/ml ampicillin and grown at 37°C until the OD<sub>600</sub> reached 0.5. Protein expression was then induced by addition of 0.4mM IPTG for 3 hrs at 37°C. The induced bacterial pellets were centrifuged at 4K X 15 min at 0°C, washed twice in GST-lysis buffer (50mM Hepes Tris.HCl pH 7.9, 0.15M NaCl, 10% glycerol, 1mM EDTA), repelleted and either immediately lysed or stored at -80°C. Bacterial pellets were resuspended in 20 mls 50mM HEPES pH 7.9, 10% glycerol, 0.15M NaCl, 1mM EDTA, 1mM PMSF, 10mM β-mercaptoethanol, 1% Triton-X-100 and the protease inhibitors leupeptin, antipain, chymostatin, pepstatin A and aprotin were included at 5µg/ml to minimise proteolysis. The cells were then lysed on ice by mild sonication in 0.1% Triton-X-100. Samples were centrifuged at 16 000 RPM at 0°C for 30 min. 500µl of GSH-Sepharose beads, (Sigma G-4510) were washed with lysis buffer and incubated with the extract for 2 hr at 4°C. The GSH-Sepharose beads were washed three times with lysis buffer and finally the bound proteins were eluted by addition of 2mls lysis buffer containing 15mM reduced glutathione (rGSH). Aliquots from these fractions were analysed by SDS-PAGE and stored at -80°C.

## **2.13 DNA binding assays**

### **2.13.1 Electrophoretic mobility shift assay**

Bandshift binding reactions were carried out in 20mM HEPES pH7.9, 1mM EDTA, 10mM  $\beta$ -mercaptoethanol, 0.1% Triton X100 and 4% glycerol. Methylated and unmethylated probes were end-labelled with [ $\alpha$ - $^{32}$ P]-dCTP using the Klenow fragment of DNA polymerase I. Aliquots of labelled probes were quantitated using a scintillation counter and 10000 cpm of each probe were incubated with recombinant proteins and poly (dI-dC) non-specific competitor on ice then loaded on a 5% polyacrylamide gel. Alternatively, samples were loaded on a 1.5% agarose gel in 0.5X TBE and DNA-complexes resolved by running at 10V/cm at 4 $^{\circ}$ C for 3 -5 hours. Gels were dried down onto DE81 filter paper and exposed to XAR5 film (Kodak).

### **2.13.2 Southwestern assay**

Proteins were denatured in 1X SDS/sample buffer (2X: 62.5mM Tris.HCl pH6.8, 2% (w/v) SDS, 0.04%  $\beta$ -mercaptoethanol, 10% glycerol, 0.1% bromophenol blue) at 100 $^{\circ}$ C for 5 min then loaded onto a 10% polyacrylamide gel and subjected to SDS-PAGE under constant amperage. Polyacrylamide gels, nitrocellulose filters (BA85, Schleicher and Schuell) and Whatman 3MM were then soaked in 1X transfer buffer (10X transfer buffer : 30.28g Tris base, 142.63g glycine per litre) for 30 min. Gels were then transferred onto nitrocellulose between 3MM filter paper and proteins were transferred in 1X transfer buffer by electroblotting at 350mA for 4 hrs. Following transfer, the filter was denatured by washing with 6M guanidine

hydrochloride in 1X transfer buffer for 30 min at room temperature. Proteins immobilised onto nitrocellulose were then renatured by washing with four successive two fold serial dilutions of the denaturation buffer using 1X transfer buffer. Filters were then washed in 1X binding buffer (10X binding buffer : 200mM HEPES pH 7.9, 30mM MgCl<sub>2</sub>, 400mM KCl) and β-mercaptoethanol was added to a final concentration of 10mM. Membranes were blocked in 2% non-fat dried milk (Marvel) for 30 mins. Filters were then rinsed in 1X binding buffer to remove excess blocking buffer and the nitrocellulose was incubated in pre-incubation buffer (1X binding buffer, 0.1% Triton X 100, 10mM β-mercaptoethanol and *E. coli* non-specific competitor) for 10min before addition of 150 000 cpm of radiolabelled DNA probe per ml. Filters were then washed in wash buffer (1X binding buffer : 0.01% Triton X100, 10mM β-mercaptoethanol) three times for five mins and filters air dried and exposed to XAR5 film (Kodak).

## **2.14 *Caenorhabditis elegans* handling, culture and genetic methods**

### **2.14.1 Handling of Nematodes**

The observation and genetic manipulation of nematodes was carried out on NGM plates (NG agar : 3g NaCl, 2.5g Bacto-peptone (Difco) and 17g Bacto-agar (Difco) in 975 ml distilled water. Following autoclaving, 1 ml of cholesterol in ethanol (5 mg/ml), 1ml of 1M CaCl<sub>2</sub>, 1ml of 1M MgSO<sub>4</sub> and 25ml 1M potassium phosphate buffer (pH 6.0) were added in order). Strains were maintained by transferring a single hermaphrodite using a 32-gauge platinum wire to 9cm NGM plates seeded with OP50, a uracil auxotroph of *E. coli* (Brenner, 1974), and were incubated at 15°C with cultures requiring sub-culturing every 10 days. All worm strains were derived from wild type Bristol N2 unless otherwise stated.

### **2.14.2 Isolation of *C. elegans* insertion and deletion mutants**

Thousands of worm cultures can be screened for either Tc1 transposon insertion events or EMS/TPA induced deletions by employing nested genomic PCR to analyse tens of orderly pooled DNA lysates (Zwaal et al., 1993; Jansen et al., 1997). These DNA lysates have a corresponding collection of frozen mutants to allow subsequent retrieval of the mutant animal. Typical libraries consist of three sets of 960 cultures of (frozen) worms : 10 racks of 12 X 8 and a corresponding set of crude lysates. A detailed description of the mutant libraries screened in this study are provided in sections 6.2.1 and 6.2.2.



Tc1 transposon insertion library screening : The DNA lysates described were analysed by PCR to visualise insertions of Tc1 transposons into a gene of interest. The PCR employed two sets of primers where one is specific for Tc1 and the other was specific to the gene of interest where products were obtained only if the Tc1 and gene-specific primers were properly juxtaposed. The reaction is made more specific and more sensitive by employing two rounds of PCR, using nested primers in the second round. The strain employed is MT3126 (Finney, 1987), which contains the *mut-2* mutation, as a result of which it is permissive for Tc1 transposon excision.

EMS/TPA library screening : The DNA lysates described above were analysed by PCR to visualise an EMS/TPA induced deletion within a gene of interest. The PCR employed two sets of nested oligonucleotide primers approximately 3.0kb apart 5' to *pbm-1* to facilitate the isolation of loss-of-function or reduction-of-function alleles. The reaction was made more specific and more sensitive by employing two rounds of PCR, using nested primers in the second round. When an address was identified, frozen oligoclonal stocks were thawed and plated onto a seeded 10cm NGM plate. Three days after thawing the culture, ~500 individual young adults were picked and transferred to OP50 seeded 9cm NGM plates and at the same time oligoclonal cultures of rising complexity were seeded by establishing 10 plates of 10 worms, 5 of 20 worms, 5 of 50 worms and a single plate of the remaining worms. Two to four days later, the parental worms that have laid viable eggs on plates and contain viable progeny were picked into 2.5µl of single-worm lysis buffer (50mM KCl, 10mM

Tris-HCl pH 8.3, 2.5 mM MgCl<sub>2</sub>, 0.45% Nonidet P-40, 0.45% Tween 20, 0.01% gelatin, freshly added 60µg/ml proteinase K) and the lysis was carried out at 65°C for 60 min followed by a protease heat inactivation at 95°C for 10 min. PCR genotyping was then performed using the outer PCR primers on these single worm DNA lysates and as a positive control a microliter of the original DNA sample was used. Nested PCR was performed on approximately 15ng of genomic DNA in a 9600 thermocycler (Perkin-Elmer) and Taq polymerase (BRL) with 0.5 µM of each primer, 0.2mM each dNTP for 35 cycles of 95°C X 40s, 58°C X 40s and 72°C X 1 min. Following identification of the mutant in this series, single progeny of the mutant line were PCR genotyped (with a negative neighbour as a control) to determine whether a mutant is homozygous. TA-cloning of the deletion mutant PCR product and subsequent DNA sequencing (section 2.8) was carried out to identify the nature of the deletion. Internal primers to the deletion were then designed to facilitate PCR genotyping to determine whether lines are heterozygous or homozygous for the deletion.

### **2.14.3 Outcrossing of *C. elegans* deletion mutants**

Five adult hermaphrodites homozygous for *dpy-17 (e164)* and *unc-32 (e169)* III (SP471) were mated with 15 young adult males *him-8 (e1489)* IV on a seeded NGM plates at 20°C. 15 male *cross*-progeny i.e. from plates that contained phenotypically wild type worms as well as DpyUnc worms, were mated from the third outcross stage with adult hermaphrodites homozygous for the *pbm-1* deletion. Twenty L4 hermaphrodite progeny from this cross were transferred individually to NGM plates

and allowed to self-fertilise. F<sub>2</sub> progeny from lines heterozygous for the *pbm-1* deletion by PCR genotyping (as described in sections 2.14.4 using oligonucleotide primers presented in sections 6.2.1 and 6.2.3) were allowed to self, PCR genotyped and the lines that were positive for the deletion were identified. This scheme was repeated five times and a homozygous *pbm-1* strain was established after three outcrosses suggesting the removal of a linked mutation that was recessive lethal. The five times outcrossed strain was deposited with the Caenorhabditis Genetics Center (CGC, University of Minnesota, USA) as *pbm-1 (qa1700)*.

Outcrosses one and two were performed by mating hermaphrodites heterozygous for the *pbm-1* deletion with *dpy-17;unc-32* heterozygous males and following the segregation of the deletion allele by PCR genotyping of their progeny.

#### **2.14.4 PCR genotyping of nematodes**

Single animals were transferred to 10µl single-worm lysis buffer (see 2.14.2) and lysis was carried out at 65°C for 60 min, followed by a protease heat inactivation at 95°C for 10 min. PCR genotyping was then performed using the outer PCR primers on these single worm DNA lysates. Nested PCR was performed on approximately 15ng of genomic DNA in a 9600 thermocycler (Perkin-Elmer) and Taq polymerase (BRL) with 0.5µM of each primer, 0.2mM each dNTP for 35 cycles of 95°C X 40s, 58°C X 40s and 72°C X 1 min.

# **CHAPTER THREE: IDENTIFICATION OF A FAMILY OF PHD-ZINC FINGER, BROMODOMAIN AND METHYL CpG-BINDING DOMAIN (PBM)-CONTAINING PROTEINS**

### 3.1 Introduction

The sequencing of the genome of the small nematode, *Caenorhabditis elegans* is now complete (The *C. elegans* sequencing consortium, 1998). This represents the first example of the entire nucleotide sequence of a multicellular organism and 19 099 genes are evident within the *C. elegans* genome. The nematode worm, *C. elegans* is also understood at an anatomical level in great detail with a complete description of the cell lineages that gives rise to the 959 somatic cells of the adult hermaphrodite nematode existing (Sulston et al., 1983) and an exhaustive description of the 302 neurons elucidated down to the level of individual synapses (White et al., 1987).

Simpson et al., (1986) were unable to detect 5-methylcytosine in the genome of *Caenorhabditis elegans* at any stage of development and as a result it was considered to be unlikely to identify any of the molecular machinery associated with a methylation system such as DNA methyltransferases and methyl-CpG binding proteins.

### 3.2 Results

#### 3.2.1 Identification of the MBD-containing gene, *pbm-1* in the nematode *Caenorhabditis elegans*

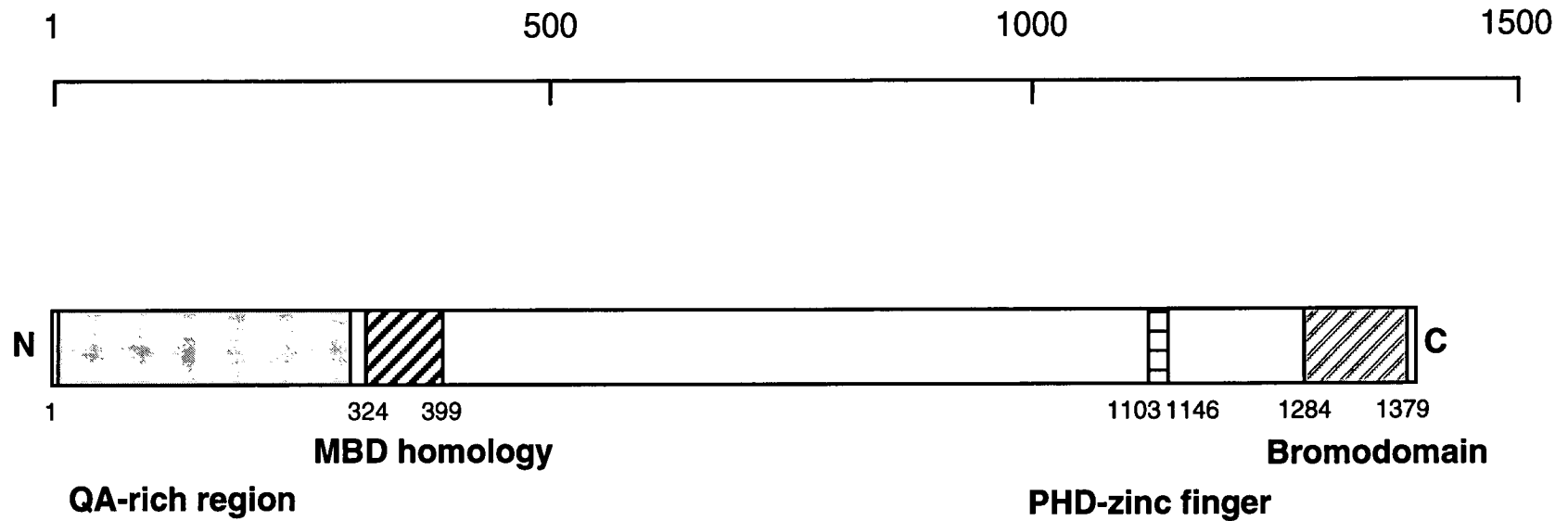
Following the screening of protein databases with the MBD of MeCP2 to identify candidate methyl-CpG binding proteins, a MBD domain was identified within a putative *Caenorhabditis elegans* gene (B. Hendrich, unpublished observations). Cosmid ZK783 (Accession No. U13646) from *C. elegans* chromosome III was

predicted to contain a gene (ZK783.4), using GENEFINDER exon-prediction software (Green and Hillier, unpublished) that potentially encoded a large protein containing an N-terminal MBD and an extreme C-terminal PHD-zinc finger region and adjacent bromodomain (see Figure 3.1). The gene was given the standard name *pbm-1* for PHD-zinc finger, Bromodomain and MBD gene-1. The MBD of *C. elegans* PBM-1 was found to be highly homologous to a human EST (593380) from an ovarian cancer cell line suggesting evolutionary and functional conservation of this domain. A project was undertaken to isolate and characterise these genes using a cross-species approach exploiting the advantages particular to the nematode and mammalian systems.

### **3.2.2 Characterisation of the genomic structure and cDNA encoding PBM-1**

A cluster of *C. elegans* ESTs derived from the *pbm-1* locus was identified by searching dbEST (NCBI BLAST server: <http://www.ncbi.nlm.nih.gov/BLAST>) and XREF (<http://www.ncbi.nlm.nih.gov/XREFdb>) with the predicted full length PBM-1 protein. The largest clone, yke2 was sequenced entirely to confirm identity between this EST and the predicted cDNA from the *pbm-1* locus determined by using the GENEFINDER exon-prediction software to analyse the genomic sequence (Wilson et al., 1994; Green and Hillier, unpublished software).

The yke2 clone however, did not extend far enough 5' to include all of the predicted *pbm-1* exons 1 and 2. In order to derive a complete 5'-end cDNA, a spliced leader



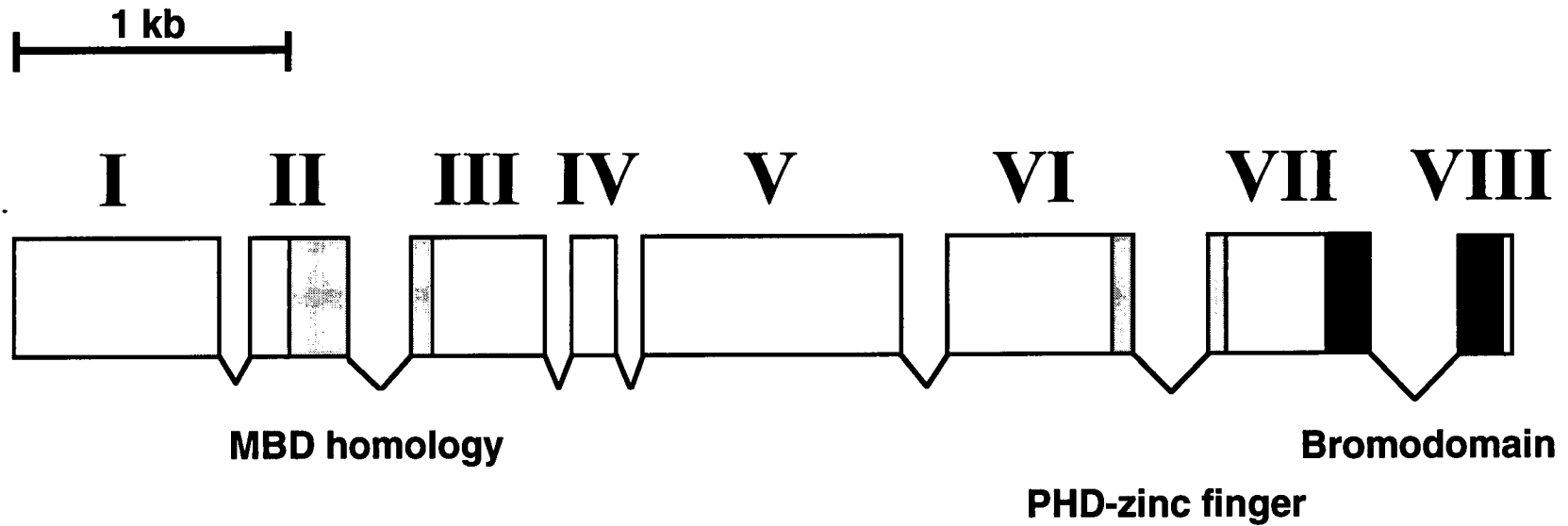
**Figure 3.1** : Schematic representation of the *C. elegans* PBM-1 protein

The locations of the glutamine-alanine (QA)-rich region, methyl-CpG binding domain (MBD), Plant homeodomain (PHD)-zinc finger and bromodomain motifs within the PBM-1 protein are indicated. The scale bar at the top of the diagram corresponds to amino acid positions numbered from the N-terminus.

primer (SL1 : 5'-GGTTTAATTACCCAAGTTTGAG-3'; Krause et al., 1989; Spieth et al., 1993) and a gene-specific primer derived from *pbm-1* exon 3 (zk783p1 : 5'-CCATTGGATGTGGACGCCTG-3') were used in an RT-PCR reaction to amplify from total *C. elegans* cDNA the intervening sequence. This PCR product was TA-cloned, sequenced and the clone was termed Cepcm10 (S.Tweedie, unpublished results). This method takes advantage of the fact that approximately 70% of all *C. elegans* mRNA messages are *trans*-spliced with either a SL1 or SL2 derived spliced leader sequence (Spieth et al., 1993; Hengartner and Horvitz, 1994; Hope, 1994) and is a commonly used method to demonstrate that a message is in fact *trans*-spliced. Thus, the *pbm-1* gene encodes a *trans*-spliced SL1 mRNA.

A cDNA contig for *pbm-1* was then derived by aligning the Cepcm10 and yke2 clones and comparing this with the predicted cDNA product from the genomic sequence of cosmid ZK783. This procedure was applied to exclude the possibility of any errors arising from failure of the GENEFINDER exon-prediction software to accurately assign intron-exon boundaries within the *pbm-1* gene. The original submission by the *C. elegans* sequencing consortium of the genomic structure of the *pbm-1* gene suggested it had 12 exons on the basis of non-experimental computer analysis. The sequence of the complete PBM-1 cDNA (i.e. Cepcm10/yke2 described above) and alignment with the *pbm-1* genomic locus demonstrated that the locus has in fact eight exons interspersed with small introns. The corrected genomic structure for *pbm-1* and the intron/exon boundaries are presented in Figure 3.2.





**Figure 3.2 :** Genomic structure of the *C. elegans pbm-1* gene

Exons are numbered in Roman numerals and the sizes of exons and introns are drawn to scale. The positions of the MBD, the PHD-zinc finger and the bromodomain motifs within the *pbm-1* genomic locus are indicated.

Small introns are typical for most *C. elegans* genes that are, on average, smaller than other eukaryotic introns (Blumenthal and Thomas, 1988). The introns within *pbm-1* conform to the “GT/AG” rule in *C. elegans*, and other eukaryotes where the 5’ splice donor sequences fit the consensus **GTRART** (R=purine) and the 3’ splice acceptor sequences terminate with the pyrimidine rich consensus **TTTCAG** (see Table 3.1; Fields, 1990).

No systematic study exists for defining an optimal consensus for translational initiation in the nematode system. Krause (1995) has presented a preliminary compilation of translational initiation sites for 48 *C. elegans* mRNAs and defined the degenerate consensus sequence  $ANNA^A/C^A/GAAAUGN$ . All of these 48 proteins in *C. elegans* are initiated at AUG and this initiator codon was frequently preceded by A residues but little similarity exists to the Kozak consensus sequence for translational initiation of vertebrate mRNAs. The PBM-1 cDNA sequence is 4173 nucleotides in length and contains a long open reading frame (ORF), initiated by the putative ATG at position 50 and terminated by the STOP codon (TAA) at position 4220 (see Figure 3.3 for cDNA and amino acid sequences). The ATG at position 50 was chosen as the initiator methionine for conceptual translation of the PBM-1 cDNA because it resides within an excellent context for translational initiation when compared to the nematode consensus sequence and is preceded by numerous A residues as has been found for a number of *C. elegans* mRNAs (Krause, 1995; 10/11 matches: UCCAAAAATGA). Furthermore, an upstream 5’UTR stop codon defines the most

<u>3' Splice Acceptor</u>		cDNA nucleotide position	Exon Number	Exon Length (bp)	<u>5' Splice Donor</u>		Intron Number	Intron Length (bp)
Intron	Exon				Exon	Intron		
	NA	1-744	1	744	AAGGCG/ <b>gtaaga...</b> K A		1	52
...tttcag/GATCAA D Q		744-1119	2	375	GTTAGG/ <b>gtaagt...</b> V R		2	175
...tttcag/TACCTC Y L		1120-1639	3	520	GGAATAG/ <b>gtaagt...</b> G I		3	48
...tttcag/ATCTTGAA D L E		1640-1778	4	139	GAGAAAAG/ <b>gtaatt...</b> E K R		4	47
...tttcag/ATTTGGC F G		1779-2714	5	936	CAGAATAG/ <b>gtcagt...</b> Q K R		5	131
...tttcag/AGACTGT D C		2715-3390	6	676	TTTAGG/ <b>gtgagt...</b> F R		6	269
...tttcag/CCCCGA P R		3391-3992	7	602	AAACTTAT/ <b>gtaagt...</b> K L I		7	296
...ttccag/ATACGAG Y E		3993-4432	8	181	NA		NA	NA

**Table 3.1 :** Exon-Intron boundaries in the *C. elegans pbm-1* gene

Exon sequences are shown in uppercase and written as codons with the translated amino acid below (single-letter code) . Intron sequences are shown in lower case. The conserved GT 5' splice donor and AG 3' splice acceptor sequences are written in bold. The cDNA nucleotide positions are numbered from the A of the ATG initiator methionine shown in Figure 3.3. NA, not applicable.

5' limit of the *C. elegans* PBM-1 cDNA and the ATG at position 50 represents the first in-frame methionine codon.

The maturation of the 3' end of a *C. elegans* primary mRNA transcript appears to use an analogous pathway to that in other eukaryotes involving cleavage and polyadenylation catalysed by an enzymatic complex that recognises the poly(A) signal sequence, AAUAAA and a second undefined downstream element (Irniger and Braus, 1994). An analysis of some 1300 *C. elegans* cDNA clones showed that polyadenylation usually begins normally 13 nucleotides downstream of a sequence matching the previously defined poly(A) signal sequence (T. Blumenthal, O. White and C. Fields, unpublished observation). Variations on the poly(A) hexanucleotide signal sequence in *C. elegans* normally involve the first position (22% of transcripts have other bases than A) and the fourth position (13% had G) as is in the case for the PBM-1 poly(A) signal sequence. In the PBM-1 cDNA sequence 3'-UTR a AAUGAA poly(A) signal sequence is evident 14 nucleotides upstream of the poly(A) tail sequence (at cDNA positions 4467-4472). The *C. elegans* PBM-1 cDNA sequence potentially encodes a large protein of 1390 amino acids with a predicted molecular weight of 156kDa. No paralogs of the *C. elegans pbm-1* gene are evident within the worm genome and furthermore, no PHD-zinc finger, bromodomain and MBD-containing proteins were identified in the completely sequenced genome of *Saccharomyces cerevisiae* suggesting a relatively recent evolutionary origin in eukaryotes.

### Figure 3.3 : cDNA and amino acid sequence of *C. elegans* PBM-1

The PBM-1 cDNA sequence is 4173 nucleotides in length and contains a long open reading frame (ORF), initiated by the putative ATG shown in bold at position 50 and terminated by the STOP codon (TAA) at position 4220. The 22 nucleotide 5' *trans*-spliced leader (SL1) is shown in bold. A polyadenylation signal AATGAA 14 nucleotides upstream of the poly(A) tail is underlined. The deduced peptide sequence of PBM-1 is 1390 amino acids in length.

1	GGTTTAATTACCCAAGTTTGAGAATTGTACATGTAAATCTTTCCAAAAATGAGTGATAACTCATCTAATCAGTTTCTACTTCTGCTAGCAGCTGCACAG	100
	M S D N S S N Q F L L L L A A A Q	
101	CAACAACAGCAACAACAACCTTCTACAGCAACAACCTTGCAAAGATCCAGAAGGCCACTGCATCATCTCCATCGAAATCAACTAATGGAACATCAGCATCAA	200
	Q Q Q Q Q Q L L Q Q Q L A K I Q K A T A S S P S K S T N G T S A S T	
201	CGTCGGCTGTCCCGTCAACTTCTGGGACATCGTCATCTCAAAATGAAGCAGCTCAGTTGCAAAATTTAGCAAAAATGCAGCAAATTCAGCAGCTGGCCCA	300
	S A V P S T S G T S S S Q N E A A Q L Q N L A K M Q Q I Q Q L A Q	
301	GTTTGAGCACTCATGGCTGCGCAGAAGAAACAACAAGAAAAGCGGCAGCTGATAAGGCAAAGGAAAAAGAAAAGGAGAAGCAAAGGCTGCAGCAGCA	400
	F G A L M A A Q K K Q Q E K A A A D K A K E K E K E K Q K A A A A	
401	GCTGCCGCCGCCGAGCAAAAGCCTCAGCGTCTACTTCCCTCCGCATCAGCAATTCCTGGATTGTCTCCAGAAATGCTCGCAGCATGGCAACAAGCAATCC	500
	A A A A A A K A S A S T S S A S A I P G L S P E M L A A W Q Q A I Q	
501	AGATGCAAGCTCTCCAACAATGATGATGACTCCACAGAAATCTCAAATGGAAGAAGCTATCAAGAAAATGATGGATATGGCAAAGAAGAAGCCAGCTGG	600
	M Q A L Q Q M M M T P Q K S Q M E E A I K K M M D M A K K K P A G	
601	AGTCGCCTCAACAAGTTCAGCATCTACAAGCTCATCCACACCATCCACATCATCAGCTTCAATTACATCATCTAATAACAATGCAGCGAACAACGCTGCG	700
	V A S T S S A S T S S S T P S T S S A S I T S S N N N A A N N A A	
701	AGCAATATGATGAACAATGTGATGTGGCAATTGGTGGCAGCACAAATGCAACAAAACAGCAACAACAAGGATACTCAGAAAAAGGCGGATCAAG	800
	S N M M N N V M W Q L V A A Q M Q Q K Q Q Q Q Q K D T Q K K A D Q A	
801	CGAAAAAGCGAAGAGCTTGCCAAACAACAAGAGCAAGATGTGAAGAATAAGCAACAAGAGGAGATTTTGAAGTTTCTTATGGCTCAACATCA	900
	K K A K E L A K Q Q Q K E Q D V K N K Q Q E E I L K F L M A Q H Q	
901	GCTGAATCATCAGAAAAAGCATGAGAAGAAGCAAGCTGACGCAGCTGCACTCGCAGCAAAAGTGCTTGCAAGCTCACCGAGCAGCACTCGAATCGGACAGT	1000
	L N H Q K K H E K K Q A D A A A L A A K V L A A H R A A L E S D S	
1001	CCTGAAGAAGGGAAAAAGACTAATGAAGCAATGCTTCGATTGCCTCTTCAACTTGGATGGAGGCGTCAAACATGTGTGAGAAGTATTGCTTCCGCTGGAG	1100
	P E E G K K T N E A M L R L P L Q L G W R R Q T C V R S I A S A G V	
1101	TTAAAGGAGATGTTTCGTATTTTCTCCGTGCGGAAAAAGTTGAGCACATACTCGGAAGTTGTTAGGTACCTCACCAAAAATTCGATTCACTACATCAC	1200
	K G D V S Y F A P C G K K L S T Y S E V V R Y L T K N S I H Y I T	
1201	CCGTGACAATTTTCTCTTCAACACAAAATGGTTATCGGGGAATTCATTGTTCCAAAGCAAACAGAAGCTGATGAGACTCAACAAGAACGCGAATTTGCA	1300
	R D N F L F N T K L V I G E F I V P K Q T E A D E T Q Q E R E F A	

1301	ATGTTTACTGAAGATGACATTAAACAAGGAACCTACAAGATTAAATGTTCTGAAATTCGTTCCAAAGATCCAGGCGTCCACATCCAATGGTGTTCACGAGG	1400
	M F T E D D I N K E L T R L N V L K F V P K I Q A S T S N G V H E D	
1401	ATGATATCAAATGTCCAAATGAAGAACCCGACGAGCCATTGGATCCATCGGAGCTTAACGATGAATTCACCGAAGAGTTAGTTCATAGTCAGATTAT	1500
	D I K M S K I E E P D E P L D P S E L N D E F T E E L V H S Q I M	
1501	GTCAAATGGAGTTGATGAATGCAAGATTCGGGAGAGAGAAGCTGATGATCTACTAGTAAATATCAATGATGTTTCGGCATCTCCAGATTTTCTAGAATT	1600
	S N G V D E C K I R E R E A D D L L V N I N D V R H L P D F S R I	
1601	GGTAATCAATGTCTGAGTTCACAAGGATTCGCTGATGCACTTATGGTTCACGAATTTGTACAAAACCTTTGGTCATGTTTtaggaATAGATCTTGAAATCG	1700
	G N Q C L S S Q G F A D A L M V H E F V Q N F G H V L G I D L E I A	
1701	CCCCAAAACCTAGAGTCGTTATGCGCAGGATTAGATGGAGATGCTAATCACGCCGAGCAAACCTCTTCAATTAACCTAGGCAACTATTAAGACTAGCACTTGA	1800
	P K L E S L C A G L D G D A N H A E Q T L Q L T R Q L L R L A L E	
1801	ATTCCTGGAATGGGAAATGAGAAAAGATTGGCCAAGGTGGAGGAGAAATGGGACTTGATCGTGAAAACCTTCAGTGAAGTGATGCGTCTTTTCTTATC	1900
	F P G M G N E K R F G Q G G G E M G L D R E N F S E V M R L F L I	
1901	GACAAAGGAAAGCGAGGAGAAGATTATCACAGCCTCTGCTTACTTGCAACTTTTTATCGATTTCTCCAGAACAAAAGCTTCGATTTTAGCATTTCTCT	2000
	D K G K R G E E L S Q P L L T C N F L S I S P E Q K A S I L A F L C	
2001	GCGATGAGCTAGTTTGTCTCGAAATGTTGTCACTGAGATTGATAAGAACCTTGATGAAATTTCAAGATTAAGGAGAAAAATGGATGAGAGAGGGAAA	2100
	D E L V C S R N V V T E I D K N L D E I S R L K G E K W M R E G K	
2101	AGCTAGAGCTTTGAGATCAGCAAGAAGTAAAAAGAAGAATGACGAGAAAGTTGTAGTTGTGAAAGAAGAACAGAATCATGAGAGTGATTCTGAACCTCCA	2200
	A R A L R S A R S K K K N D E K V V V V K E E Q N H E S D S E P P	
2201	ACTAGACCGGATACTCCGAAGAAAGCCACGGTTGCACCTCCAACCTGTAGTTTCTGTTTACCTGTCTCGGCAGCTCAGCAGCAACAACGAAAGTTCACCTC	2300
	T R P D T P K K A T V A P P T V V S V S P V S A A Q Q Q Q R K F T P	
2301	CAGGATTAGGCCAGTGCAGAGTTTAAACAGAACAAGAAGATCAATGAGTCTTCAACAAATGGATTCTCTGATCGGAGATCTTCATCAAGAAGCTCAAAA	2400
	G L G Q C E V L T E Q E E S M S L Q Q M D S L I G D L H Q E A Q N	
2401	CATCAATCAGAAGATTCATGATACTGGTCTCAAGATCCGATCCTTCCCATTCCGAACTGACCGTTTTCATCGAAACTACTGGATGTTGGCTCACACTGAT	2500
	I N Q K I H D T G L K I R S F P F G T D R F H R N Y W M L A H T D	
2501	AAAGTTATCATTGAGTCACTAGCAACAACCTTCAGTGAACAACCCAGCCTGCAATGCAACGAATATGCATCAAAGGATCCACCAACTCTTGAACAGCGTG	2600
	K V I I E S L A T T S V N N P A C N A N E Y A S K D P P T L E Q R V	

2601	TTCCAGGAGCATGTGAGACGATTGACTTAGATGTTATTGCTTGTGTTGAGGATTGGTTGATGATGTTGTTCTGCTAAGAGCCAAAGCTGATAAGAAAAC	2700
	P G A C E T I D L D V I A C V E D L V D D V V L L R A K A D K K T	
2701	TCGAAAAAGATACAGAAGGATTGAAAATCATATGAAGAGAGGATGGTGGAATGCAGAATAGAGACTGTGTAGAATCCCTTCGTTCTTGTATGCTCAGT	2800
	R K R Y R R I E N H M K R G W W T M Q N R D C V E S L R S C M L S	
2801	CGCGGAATTCGTGAACGTGCTCTTCATCGTCTACTCACTAAACCATGGTTTCTAAATGAACTCAAATTCGGTACAATCACAATCGAACCAGTTGGTGAGA	2900
	R G I R E R A L H R L L T K P W F L N E L K F G T I T I E P V G E K	
2901	AATCCGACCTGGAACGGTTCGAAGACAAGGATGGACTCGACTAAATACGGCTATCGACAAGCTTCAATGTCATTTGAAAATGTCAGATGTGTCTAAACC	3000
	S D L E L V R R Q G W T R L N T A I D K L Q C H L K M S D V S K P	
3001	ATTGCCTTCTATCACGCCTTTCGAGACCCAGAAGCCTATTGTGGTGCCTCCTACAATGGCATTGGCGCAAATTTGTGAAGGACGATATGGCATGGAAGGTT	3100
	L P S I T P F E T Q K P I V V P P T M A L A Q I V K D D M A W K V	
3101	ATAGATGAGGAGTTGATGGACAAGAACTCGATGAGACTATAATCCGACAAAAAATCATTGAAACAGCGGACATGGTGCAGCCAAAGTTTGGAGACCCA	3200
	I D E E V D G Q E L D E T I I R Q K I I E T A D M V Q P K F W R P K	
3201	AATTTCAAAACTGGAGGATCAAGATACATGCCAACTATTTGAGGATTGGAAGTCTTATGTTTCAACTGAAGCACAAACAACGAGTCAGTTAATGGTAGC	3300
	F Q K L E D Q D T C Q L F E D W K S Y V S T E A Q T T S Q L M V A	
3301	TCTTCAAACTCTCGAAGGAATGATTATGTGGGAAAGATCATCTCGTGAAGCACTTTGTCAGATTTGTAAGAGTATGGATGGAGATGAAATGTTGGTTTGT	3400
	L Q T L E G M I M W E R S S R E A L C Q I C K S M D G D E M L V C	
3401	GATGGATGTGAGAGTGGATGCCATATGGAATGTTTTAGGCCCGAATGACAAAAGTACCTGAAGGTGATTGGTTCTGTCAGCGGTGCCGCGAGGAAAAAT	3500
	D G C E S G C H M E C F R P R M T K V P E G D W F C Q R C R E E K S	
3501	CTGGAAGACCAATGTGCATGTTCTGTAGCCGTGAGACTGGAAATCTTCATCAGTGTCAACGCTGTGCATACCATGTTTCATCAAGAATGTTCAAGATGG	3600
	G R P M C M F C S R E T G N L H Q C Q R C A Y H V H Q E C S Q D G	
3601	CCCCAAAGAGGCTATAAACCTGAAACATTTATCTGTGGGCATTGTCAAGAAATGAAGCAAATGCGTTTTGTGAAGCGACTCATTCTTCGAAGTGAATCG	3700
	P K E A I N P E T F I C G H C Q E M K Q M R F V K R L I L R S E S	
3701	GAAGAACGAGAATTGGAAGATGATAATCATGCAGAAAATGGAGAGAATACTAAAAATGGTCATATGAATGGAATGAATGGAGCAATTGCAATCGGCGTTC	3800
	E E R E L E D D N H A E N G E N T K N G H M N G M N G A I A I G V H	
3801	ATAATCAACAAAACGGAGTCAAAGGAAATCTGAAGAGAAAATTGGAGGTTCCAAGTATTGGTGGTCTTCCAAAAATATGAATAAGGAGTTGTGTCAACT	3900
	N Q Q N G V K G N L K R K L E V P S I G G L P K N M N K E L C Q L	



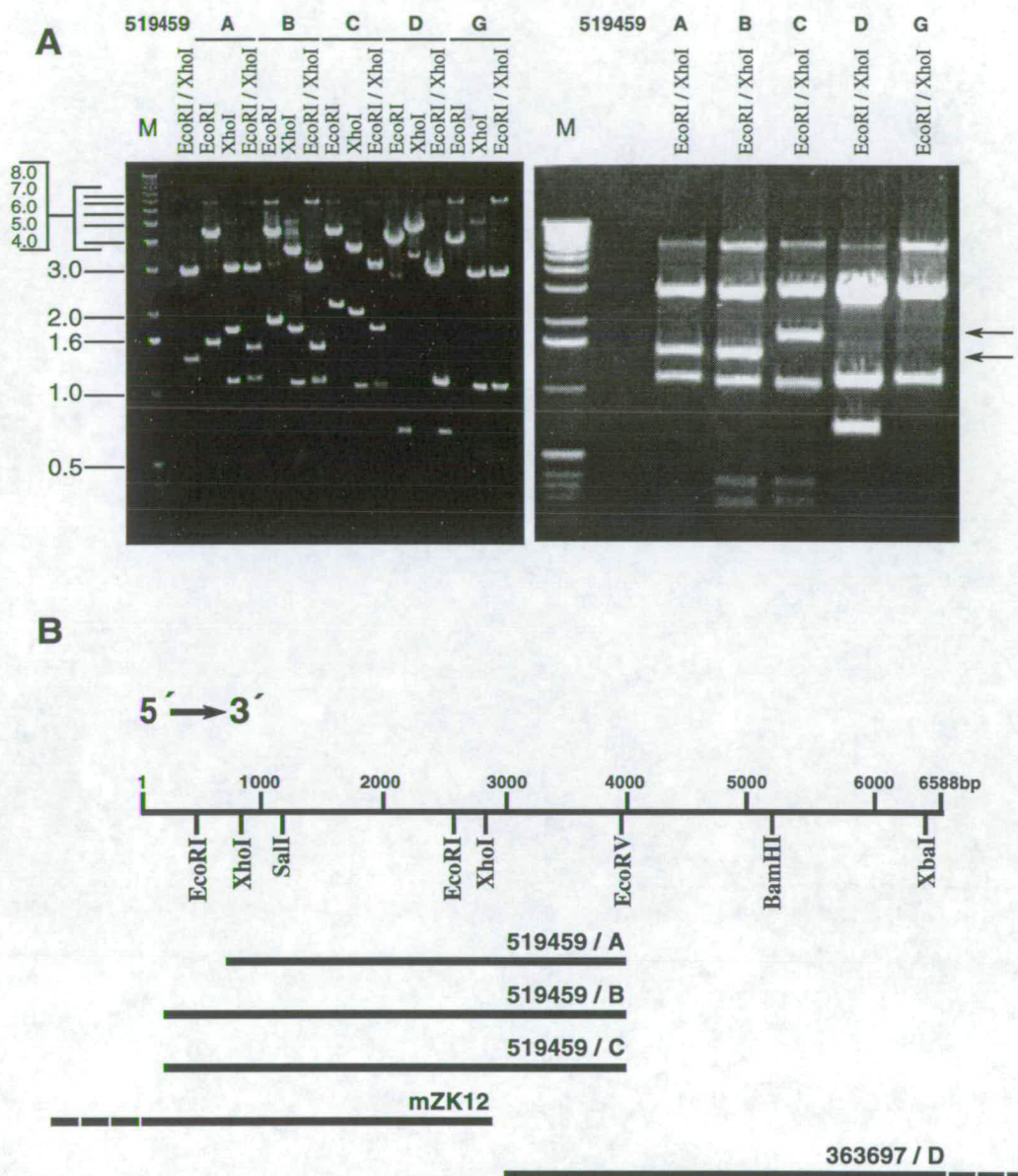
3901	CATGTTGGATGAACTCGTCGTCCAAGCAAATGCTTTGCCATTTTGGAGCCTGTCAATCCAAAATTGGTTCCTGGATATAAAATGATTATTTCAAACCA	4000
	M L D E L V V Q A N A L P F L E P V N P K L V P G Y K M I I S K P	
4001	ATGGACTTGAAGACAATTCGACAGAAGAATGAGAACTTATATACGAGACTCCAGAAGATTTTGCTGAAGACATAGAGCTCATGTTCCGCAACTGTCGTC	4100
	M D L K T I R Q K N E K L I Y E T P E D F A E D I E L M F A N C R Q	
4101	AATCAACATTGATCACTCAGAAATTGGACGTGCTGGAATCAGTTTGCACAAGTTTTTCCAAAACGTTGGAAGCAGCTGAAATACAACCTTACAAAAG	4200
	F N I D H S E I G R A G I S L H K F F Q K R W K Q L K Y N F T K R	
4201	ACTCAGACGCCTTCATCGTTAATTTTCATGTTTTAATCAATTATCATCTCACCGTAACCGTTAACTTTTTCTGTTCATAAATACAAAATCATCCATTG	4300
	L R R L H R *	
4301	ATGTTCGGGTAGTTATTTAATGTTTTTATAATTTGCATATTTGTGCCACATGTACAGTTGTACATTTTCTCTTTTTTAATTTATTTATCAATACCCCA	4400
4401	TCACATTTCTTTCCGTGACGGATTTCTCATGCTTTGATCTGATTATTTATTGCAATCTTGAAGAAAATGAAAATTTTCAAACAAAAAAAAAAAAA	4500
4501	AAAAAAAAAAAAAAAAAAAA 4518	

### 3.2.3 Isolation of mPBM1 - a mammalian ortholog of *C. elegans* PBM-1

*Caenorhabditis elegans* represents a powerful model system for the elucidation of various biological phenomena by the means of molecular genetics (Wood, 1988; Epstein and Shakes, 1995). The nematode worm has proven to be particularly useful for the isolation of loss-of-function mutants or hypomorphic alleles to aid in the elucidation of gene function (Plasterk, 1995) and powerful forward and reverse genetics exists in this organism. However, a serious limitation of the worm system is that no tissue culture cell lines have been established and it is problematic to isolate any homogenous cell type (Krause, 1995). Thus in order to exploit the molecular approaches possible with mouse and human cell lines it was necessary to isolate a full length murine cDNA sequence encoding a protein homologous to *C. elegans* PBM-1.

In order to isolate murine cDNAs homologous to *C. elegans* PBM-1, the corrected full length PBM-1 protein was used to search the dbEST and XREF databases and a cluster of murine (and human) ESTs highly homologous to *C. elegans* PBM-1 were identified. A 1.2kb *Mus musculus* kidney cDNA clone (519459) exhibited the highest probability score  $P(N)$  of homology ( $P < 10^{-24}$ ) to a central region of the *C. elegans* PBM-1 cDNA and following sequencing and conceptual translation was found to contain a small uninterrupted ORF.

A mouse brain (pituitary)  $\lambda$ ZAPII cDNA library (a kind gift from Michael Shipston, Physiology, University of Edinburgh) was screened with this 1.2kb *Mus musculus* kidney cDNA clone (519459) in order to isolate murine cDNAs homologous to the nematode PBM-1 sequence. From 500 000 plaques seven independent positive phage clones were identified, re-plated and the screening process repeated until finally five pure phage plaques were obtained. The cDNAs present within these phage were recovered by in vivo excision of the inserts in pBluescript following co-infection of XL1-Blue *E. coli* cells with the  $\lambda$ ZAPII phage clones and a F1 helper phage (R408). Re-infection of bacterial cells with this phagemid generated double stranded DNA to produce bacterial colonies for plasmid amplification. This in vivo excision procedure possesses the distinct advantage that it circumvents the requirement for phage DNA purification and subcloning of cDNA inserts. The inserts from these five cDNA clones were released by restriction digestion with EcoRI/XhoI and size fractionated on agarose gels (see Figure 3.4A). From these five cDNA clones, 519459/B and 519459/C were the largest spanning 3.6kb. The 519459/A clone was approximately 2.9kb and clones 519459/D and 519459/G were 0.9kb and 0.7kb respectively. These clones were sequenced entirely from each end by a primer walking approach using sequence cycling using fluorescent dye terminators (Perkin-Elmer). BLAST and FASTA searches against peptide sequence databases revealed significant protein sequence identity to the *C. elegans* PBM-1 protein ( $P < 10^{-25}$ ).



**Figure 3.4 : Isolation and restriction mapping of the mPBM1 cDNA clones**

**A :** Restriction mapping of the 519459 cDNA clones.

cDNA clones isolated from a mouse brain cDNA library with the 519459 probe were restriction digested with EcoRI, XhoI and EcoRI/XhoI and size fractionated on 1% agarose gels. An alternative splicing event (+285bp) present within the 519459/C cDNA clone but absent in the 519459/A and 519459/B clones is indicated by arrows.

Molecular weight standards are shown in kilobases.

**B.** Restriction map of the mPBM1 cDNA sequence.

The cDNA clones used to assemble the mPBM1 cDNA contig are presented as lines below the restriction map. The dashed lines at the ends of the mZK12 and 363697 cDNA clones represents the 5' and 3' UTR of the mPBM1. The cDNA clones are drawn to scale.

In particular, the largest of these cDNA clones (519459/B and 519459/C) potentially encoded a methyl-CpG binding domain (MBD) motif homologous to the MBD motif of the *C. elegans* PBM-1. However, this cDNA sequence was considered incomplete on the basis of a number of considerations. Firstly, conceptual translation of the largest of these cDNA clones generated an uninterrupted ORF containing no upstream, in-frame terminator codon to allow unambiguous demarcation of the 5' limit of the cDNA, raising the possibility that the initiator methionine had not yet been identified. Secondly, all the cDNA clones originated from an A-rich region which was likely the priming site bound by the oligo-dT primer employed in the construction of the cDNA library. Furthermore, these cDNA clones did not contain an upstream terminator codon or a downstream polyadenylation signal sequence, strongly suggesting that this A-rich region was not the true poly (A) tail. Finally, these clones were considered incomplete because no coding sequences for the 3' bipartite PHD-zinc finger/bromodomain characteristic of the *C. elegans* PBM-1 cDNA were present.

In order to identify more 5' and 3' clones to complete this cDNA sequence, 700bp fragments from the largest of these cDNA clones (519459/C) were employed as probes to re-screen the mouse pituitary  $\lambda$ ZAPII cDNA library. From this screen, 22 positive primary  $\lambda$ ZAPII clones were identified that hybridised strongly to the 5' and 3' fragments. These phage cDNA clones were re-plated and the screening process repeated until finally 14  $\lambda$ ZAPII pure phage plaques were obtained. The 14  $\lambda$ ZAPII

clones were in vivo excised, their cDNA inserts released by restriction digestion with EcoRI/XhoI and these products size fractionated on agarose gels. These cDNA clones were sequenced from each end by sequence cycling using fluorescent dye terminators (Perkin-Elmer). Using the sequences from the ends of these cDNA clones and the largest of the 519459 clones (519459/B and 519459/C) a cDNA contig was constructed employing DNA analysis software (DNASTar). Of these 14  $\lambda$ ZAPII cDNA clones, seven were identical, six were unique but all were contained within the original 519459 clones identified in the original screen. All were derived from the 5' end of the cDNA contig with no 3' clones isolated by the 0.7kb 3' fragment of the 519459 cDNA clone. A single cDNA clone however,  $\lambda$ ZK12 was found to contain a 3.2kb cDNA insert that extended 600bp 5' of the 519459 clones. This upstream sequence allowed the assembly of a cDNA contig termed  $\lambda$ ZK12/519459 (see Figure 3.4B). The  $\lambda$ ZK12 clone was sequenced entirely and conceptual translation of this cDNA clone generated an interrupted ORF that contained numerous upstream translation terminator codons allowing the cDNA to be defined and the assignment of a putative initiator AUG codon to this cDNA. The criteria used to assess the completeness of the cDNA sequence and specifically the choice of the initiator methionine will be described in section 3.2.4.

In order to isolate 3' cDNA clones to contig with the 3' end of the 519459 clones, the cDNA sequence encoding the *C. elegans* PBM-1 PHD-zinc finger/bromodomain region was used as a search probe to screen the dbEST and XREF databases. A

human retinal EST (363697) was identified that potentially encoded a protein containing a bipartite PHD-zinc finger/bromodomain that was highly homologous to the extreme C-terminus of the *C. elegans* PBM-1 protein ( $P < 10^{-24}$ ). A 2.2kb NotI/EcoRI restriction fragment from the 363697 human EST was employed as a heterologous probe to screen the mouse pituitary  $\lambda$ ZAPII cDNA library. From this screen, four positive primary murine  $\lambda$ ZAPII phage clones were identified that hybridised to the 363697 probe. These phage cDNA clones were re-plated and the screening process repeated until pure phage plaques were obtained. Finally, two  $\lambda$ ZAPII clones (363697/A and 363697/D) were in vivo excised, restriction digested and the products sized on agarose gels. Of the clones isolated, 363697/A contained an insert of around 3.5kb and the 363697/D clone had a larger cDNA insert of approximately 5.0kb. Using the sequences from the ends of the 363697 clones and the  $\lambda$ ZK12/519459 cDNA sequence a cDNA contig was constructed. The large 363697/D cDNA clone extended 4.0kb 3' of the 519459 clones (see Figure 3.4B). The 363697/D cDNA clone was partially sequenced by primer walking from the 5' end until a downstream TGA translation terminator codon was identified delineating the 3' end of the cDNA. The 3'UTR present within the 363697 cDNA clones was not sequenced. The  $\lambda$ ZK12/519459/363697 cDNA contig was conceptually translated and BLAST and FASTA searches against peptide sequence databases with the predicted protein sequence revealed a highly significant homology to the *C. elegans* PBM-1 protein ( $P < 10^{-79}$ ). This murine cDNA was termed mPBM1 for murine PHD-zinc finger, Bromodomain and MBD1.

The  $\lambda$ ZK12, 519459/A and 363697/D cDNA clones were ligated together to join the mPBM1 cDNA (the clones used are shown schematically in Figure 3.4B). A 3.5kb XmaI/XhoI 363697/D fragment comprising the 3' end of the mPBM1 cDNA was directionally sub-cloned into XmaI/XhoI linearised pBluescript generating a clone termed p363697/D. This p363697/D vector was XmaI linearised and a 2.1kb XmaI fragment from the 519459/A cDNA clone, comprising essentially the central region of the mPBM1 cDNA was inserted generating a clone termed p519459/363697. Correct orientation of the 2.1kb 519459/A fragment in the p363697/D vector was confirmed by restriction mapping. Finally, a 2.2kb NotI fragment from  $\lambda$ ZK12 (corresponding to the 5' end of the mPBM1 cDNA) was subcloned into NotI digested p519459/363697 generating a construct, pmPBM1. This pmPBM1 construct was restriction mapped and sequenced entirely to ensure no mutations were present within the sequence. This pmPBM1 construct is the vector employed for PCR-subcloning to expression constructs described in Chapter 4.



### Figure 3.5 : cDNA and amino acid sequence of mPBM1

The mPBM1 cDNA sequence is 7081 nucleotides in length and contains a long open reading frame (ORF), initiated by the putative ATG shown in bold at position 411 and terminated by the STOP codon (TGA) at position 6996. The presence of upstream terminator codons (\*) within the 5' UTR of the mPBM1 are indicated.

As described in section 3.2.4, the mPBM1 transcript is alternatively spliced and three major RNA isoforms of the mPBM1 transcript were detected and are termed mPBM1a, mPBM1b and mPBM1c (see Figure 3.7). mPBM1a splices out a 6bp region at the extreme 5' end of the mPBM1 cDNA (at nucleotide positions 792-798) and interestingly removes a 285bp region (positions 2360-2645) immediately 5' of the MBD. The mPBM1b isoform splices out the 6bp and 285bp sequences as in the mPBM1a transcript but also removes a 27bp region 3' of the MBD (at positions 3358-3385). Finally, the mPBM1c transcript alternatively splices a 66bp sequence adjacent to the MBD (cDNA nucleotide positions 3246-3312). The MBD of mPBM1 is shown in bold. All splice variants are underlined.

1	CATGTTTTGTGAGGGAGAGGTTGTGTTAGGTAGATTTTAGGTGTTTCTACAAACAGAGCATTGCAAATGCTTGCTTGGCTTGTGTTTTTCATTATAGCGCT	100
101	TGATAAAAGACATATTGTGACCTCTGGGGTCAGCAGAGCTAGGATACAATAGTAAGTATTCTTTCAAGCCTTTCAACAGAAAACACCGTGGCTGGCTGTG	200
201	CAGTGCTGCCTCTGCTCTGAATCGAAAGGATAATGGCTACTTCCTGATTAAAGCCGGTCCACCGTAAGCAGAGCTCCGTGCTGAATGTCTTTCTCCTGCT	300
301	AACGAGCGAGCGCGTTGCCGCTTGCTTTATTGTGTGTCTGTTTTTCATTGATTTTCTAATAGGAGCTTCTTCAGAGAATAGAAAAGAATGAGTGTAAGTT	400
	* *	
401	ATACACTGATATGGAGTCTGGAGAACTGTGCCATCCTCACCAGCCTCCTCCACTACACCAACATCATCTTCTGCTCCTCCGTGGCTTCAGCAGTTTCA	500
	M E S G E L L P S S P A S S T T P T S S S A P S V A S A V S	
501	AAGAGCAGCCTTTCCACAGGAGCTGCTTCACTTAGTTCTACAGCCAGCCCATGTGTACTTGAAGCTGGGAAGTCCAAGATCAAGGTATCACCTGACTCAG	600
	K S S L S T G A A S L S S T A S P C V L E A G K S K I K V S P D S V	
601	TGTCTGGACATTTATTTCAGAGCAGCTGGGGATCAGCCGTTTAACCTGTCCACAGTGCCGAGTGCCTTCCCAATGCTCAGCCACCCAGGCTTTGGTCTACA	700
	S G H L F R A A G D Q P F N L S T V P S A F P M L S H P G F G L H	
701	TTCAGCCAGCTCAGGGCATTTCAGAATTTGGTGGTTTGGGGACACTTGGTACACCCACAGCCTTAGCCGCACATCCCCAGCTAACATCTTTTCCAGGTGCA	800
	S A S S G H S E F G G L G T L G T P T A L A A H P Q L T S F P G A	
801	GAGTGGTGGCGAACCACAGATGGCCACTCTCGTGTGGAACGCCCTTCTTTCCACCATTACTGGGAATTCACCACTGTTTGCTCCTCCAGCCCAGAATC	900
	E W W R T T D G H S R A G T P F F P P L L G I P P L F A P P A Q N H	
901	ACGATTTCGTCATTCCATTCGAGGACTTCTGGGAAGAGCAGTCGAAATGGTCTGAGAAAGGTATAAATGGGTCGGTTAATGGAACCAAGTGCAGCATCAGT	1000
	D S S F H S R T S G K S S R N G P E K G I N G S V N G T S A A S V	
1001	ACTTGGTGTCAATGCATCTGTAGTAGCCACTCCTGCTTCAAGTTCCATGGGACAGAACCAAGCACAAGCTCAGGTGGAGGAACCCATAAATGTCATCAG	1100
	L G V N A S V V A T P A S S S M G Q N Q S T S S G G G T L K C H Q	
1101	GAACAAAACAAAAGCCAGCCTGTGGACGCCAGAGCCGACAGAATCAAAGATAAGAAACCAAGGAAGAAGGCTATGGAAAGTTCTAGCAACAGTGATAGTG	1200
	E Q N K S Q P V D A R A D R I K D K K P R K K A M E S S S N S D S D	
1201	ACTCAGGCACATCGTCAGACACCTCGAGTGAAGGCATTAGCAGCAGTGACTCAGACGATCTGGAAGAGGAAGAGGAAGAAGATCAAAGCGTGGAGGAGAG	1300
	S G T S S D T S S E G I S S S D S D D L E E E E E E D Q S V E E S	
1301	TGAAGATGACGATTCTGACTCAGAGACTGAAGCACAACACAAAAGTAACAACCAGGTGCTATTACATGGTATTTTCAGACCCAAAAACAGATGGACAGAAA	1400
	E D D D S D S E T E A Q H K S N N Q V L L H G I S D P K T D G Q K	
1401	GCAACTGAAAAAGCCCAGGAAAGAAGAACACACCAGCCATTACCTCCTGTGTCTGAGTCCCAGACGCACCCACCATTCCAGTCCCAGCAGAAGCAGCCTC	1500
	A T E K A Q E R R T H Q P L P P V S E S Q T H P P F Q S Q Q K Q P Q	

1501	AGGTTCTCTCACAGCAGCTTCCGTTTATTTTCCAAAGCTCTCAGGCAAAGGAGGAGTCTGTGACCAAACACACCAGTGTAAATACAGTCGACGGGACTGGT	1600
	V L S Q Q L P F I F Q S S Q A K E E S V T K H T S V I Q S T G L V	
1601	GTCCAATGTGAAACCTTTGTCTTTGGTAAATCAAGCCAAAAAGGAAACTTACAGGAAACTCGTAGTTCCCTTCTCCTGATGTACTTAAAGCAGGGAATAAA	1700
	S N V K P L S L V N Q A K K E T Y R K L V V P S P D V L K A G N K	
1701	AATACCTCTGAAGAGTCTAGTTCTTTGACCAGTGAATTGCGATCCAAACGGGAACAATATAAGCAGACATTCCCATCACAAGGAAAGAAACAGGAGATGG	1800
	N T S E E S S S L T S E L R S K R E Q Y K Q T F P S Q G K K Q E M G	
1801	GGAAGAGCCTGAAGAAGGTTATTGCAGCTTTGTCAAATACTAAAGCAACCTCTAGCTCACCAGCACATCCAAAACCTACCACTAGATAACAACCACCTAA	1900
	K S L K K V I A A L S N T K A T S S S P A H P K L P L D N N H P N	
1901	TCCATTTTTGACAAATGCACCTTTAGGGAATCACCAACCAATGGAGTCATTCAAAGCGTCATTCAAGAAGCTCCGTTAGCACTAACTACCAAAACGAAA	2000
	P F L T N A L L G N H Q P N G V I Q S V I Q E A P L A L T T K T K	
2001	ATGCAGAGTAAGATTAACGAAAATGTCTCAAGTAGCACCCCGTTCTCCTCACCTGTCAATCTGAGCACGAGTGGGAGGAGAGCCCTGGCAGTCAGACTC	2100
	M Q S K I N E N V S S S T P F S S P V N L S T S G R R A P G S Q T P	
2101	CAGCGCTGCCCTCTGCCTCGCCCATCCTGCACAGCAGTGGGAAGGAAAAGCGGGTCAGCAATGACGCAACCCCGCTCAAAGCGCACCATCACCCCATCC	2200
	A L P S A S P I L H S S G K E K R V S N D A T P L K A H H H P H P	
2201	GGCGGCAGCGCCCGCGCCTTGGTGGAACAGTTTAGAGGGACAGATTGAGACGTTCCAGTAGTAAAGACTCTGAGGACTCAAATGAGGATGAGGAGGAA	2300
	A A A A A A L V E Q F R G T D S D V P S S K D S E D S N E D E E E	
2301	CATGATGAGGAAGAAGATGAAGAGGATGATGAAGATGATGAGTCTGATGACAGTCAATCAGAATCAGATAGTAATTCACAATCAGATTCAGAAGGCTCAG	2400
	H D E E E D E E D D E D D E S D D S Q S E S D S N S Q S D S E G S E	
2401	<u>AAGACGACGAGGAGAAAGACCAGGAGGAATCAGATAGTGACACTGAAGCGGAGAAACCTGCCGTGAACCTGACCCAGACAAGCTCCTCTGCCAAAAGCCC</u>	2500
	D D E E K D Q E E S D S D T E G E K P A V N L T Q T S S S A K S P	
2501	<u>TCCCTCCAGTCTCACAGCGCACTCAGCACTCACCTCCACATAGGAAGCCCCCAGGCTCTGCTCCTGCTGCCCTGTGCTCTGAGTCCCAGCCTCCTGCT</u>	2600
	P S S L T A H S A P H L H I G S P P G S A P A A L C S E S Q P P A	
2601	<u>TTTCTTGGCACTTCTTCTTCCCACTCACTTCAACTCCCACTCTGGTACATCCAAACGAAGAAGAGTGGCAGACGATCAGGAGCTGCGCATCCCTCTGG</u>	2700
	F L G T S S S T L T S T P H S G T S K R R R V A D D Q E L R I P L D	
2701	ACTATGGCTGGCAGAGGGAGACAAGGGTGAGAACTTTGGTGGGCGCCTTCCAGGAGAAGTTGCATATTATGCCCCGTGTGGGAAGAAGCTGAGGCAGTG	2800
	Y G W Q R E T R V R N F G G R L P G E V A Y Y A P C G K K L R Q C	

2801	CCCTGACATGGTAAAGGGAATGCAGTGGTGTCTTTTGAAGAAGAGGATGTCATTCTCGTATCAGGGCAATGGACGGCCGGAGAGGAAGACCACCCAAT P D M V K G M Q W C L L K E E D V I P R I R A M D G R R G R P P N	2900
2901	CCAGATAGACCCAGAGCAAGAGAGGAATCCAGGATGAAACGCCGGAAGGGTCGCCCTCCCAATGTTGGCAGTGCAGAATTCCTAGATAACACAGATGCAA P D R P R A R E E S R M K R R K G R P P N V G S A E F L D N T D A K	3000
3001	AATTACTAAGGAACTGCAGGCTCAAGAAATAGCCAGACAAGCAGCACAAATAAACTTTTGAGAAAACCTCAGAAGCAGGAACAGGCACGGGTTGCCAA L L R K L Q A Q E I A R Q A A Q I K L L R K L Q K Q E Q A R V A K	3100
3101	AGAAGCTAAAAACAGCAAGCGATAATGGCTGCGGAAGAGAAGCGCAAACAGAAAGAACAGATGAAGATTATAAAGCAGCAGGAAAAAATTAAGAGAATA E A K K Q Q A I M A A E E K R K Q K E Q M K I I K Q Q E K I K R I	3200
3201	CAGCAAATCAGAATGGAGAAAGAACTGAGAGCTCAGCAGATTCTCGAGGCTAAAAAGAAAAGAAGGAAGACGGCAAATGCCAAATTATTGGAGGCCG Q Q I R M E K E L R A Q Q I L E A K K K K K E E A A N A K L L E A E	3300
3301	<u>AGAAACGAACCAAGGAGAAAGAACTGAGAAGACAGCAGGCTGTCTTCTGAAACACCAGGAGTTGGAGAGGCATAGACTAGATATGGAACGGGAGCGAAG</u> K R T K E K E L R R Q Q A V L L K H Q E L E R H R L D M E R E R R	3400
3401	GCGGCAGCATGTGATGCTTATGAAAGCTATGGAAGCTCGTAAAAAGCAGAAGAAAAGAACGATTGAAGCAAGAGAAACGTGATGAGAAAAGATTGAAT R Q H V M L M K A M E A R K K A E E K E R L K Q E K R D E K R L N	3500
3501	AAGGAACGCAAACTAGAGCAGCGGAGACTGGAGTTAGAAATGGCGAAGGAGCTAAAGAAACCTAAGGAAGACATGTGCTTAGCAGACCAAAGCCCTTGC K E R K L E Q R R L E L E M A K E L K K P K E D M C L A D Q K P L P	3600
3601	CAGAGTGGCCTCGCATCCCGGACTTGTGCTCTCTGGAACCACATTTTCAGACTGTCTCATGGTGGTGCAGTTCTTACGAACTTTGGTAAAGTTTGGG E W P R I P G L V L S G T T F S D C L M V V Q F L R N F G K V L G	3700
3701	CTTTGATGTGAATATTGATGTTCCAAACCTCAGTGTCTTCAAGAGGGATTGCTAAATATAGGGGACAGCATGGGTGAAGTACAAGACTTGCTTGTGAGG F D V N I D V P N L S V L Q E G L L N I G D S M G E V Q D L L V R	3800
3801	CTCCTCTCAGCTGCTGTATGTGATCCAGGTCTAATTACAGGATACAAGGCAAAAACAGCTCTGGGAGAACATTTGCTGAACGTTGGTGTGAATCGAGACA L L S A A V C D P G L I T G Y K A K T A L G E H L L N V G V N R D N	3900
3901	ATGTGTCCGAGGTTTTGCAGATTTTCATGGAGGCCCACTGTGGGCAGACTGAGCTGACTGAGAGCCTGAAGACTAAAGCTTTCCAGGCTCACACTCCAGC V S E V L Q I F M E A H C G Q T E L T E S L K T K A F Q A H T P A	4000
4001	ACAGAAAGCGTCCATCCTGGCCTTCCTCGTTAACGAGCTGGCATGCAGCAAGAGTGTGGTGAGTGAGATTGACAAGAACATTGAATACATGTCTAACTTG Q K A S I L A F L V N E L A C S K S V V S E I D K N I E Y M S N L	4100

4101	AGAAGAGATAAATGGATGGTGAAGGTAAACTCCGCAAGCTTAGGATAATTCATGCTAAGAAAACAGGCAAGAGAGACGCCTCAGGGGGCATTGACCTCG R R D K W M V E G K L R K L R I I H A K K T G K R D A S G G I D L G	4200
4201	GGGAAGAACAACACCCCTTGGGCACGCCCACGCCAGGACGCAAGCGGAGAAGGAAGGGAGGGGACAGTGATTATGACGATGATGACGATGATGACAGCGA E E Q H P L G T P T P G R K R R R K G G D S D Y D D D D D D S D	4300
4301	CGACCAAGCTGATGAGGATGAGGAAGATGAGGAAGATAAAGATGACAAAAAGGAAAGAAGACCGATATCTGTGAAGATGAGGATGAAGGTGACCAAACA D Q A D E D E E D E E D K D D K K G K K T D I C E D E D E G D Q T	4400
4401	GCCAGCGTTGAGGAGCTAGAAAAACAGATTGAAAACTGAGCAAACAACAGAGCCAGTACAGACGAAAGCTCTTTGATGCTTCTCACTCACTACGCTCGA A S V E E L E K Q I E K L S K Q Q S Q Y R R K L F D A S H S L R S M	4500
4501	TGATGTTTGGTCAGGATCGGTACCGACGCAGATACTGGATTCTCCCTCAATGTGGAGGGATTTTGTAGAAGGCATGGAAAGTGGTGAAGGCCTAGAAGA M F G Q D R Y R R R Y W I L P Q C G G I F V E G M E S G E G L E E	4600
4601	AATTGCAAAAAGAAAAGAAAAGCTTAAAAAGGCAGAGAGTCTCCAGATCAAAGAAGAAGTATTTGAGACTTCTGCGGAGACATTAAACTGTTCATTAGA I A K E K E K L K K A E S L Q I K E E V F E T S A E T L N C S I R	4700
4701	GATCACTGCGAGCAGAAAGACGACCCAAAAGAGAAAGATAACACAAATCTATTCTTCAGAAACCAGGCTCTTTCTCCAAGTTAAGCAAGCTTTTAGAGG D H C E Q K D D P K E K D N T N L F L Q K P G S F S K L S K L L E V	4800
4801	TAGCTAAAATGCCTCCTGAGTCAGATGTTATGACCCCCCAAAGTGAATGTTAGTACAAACGGGGGCCACTGTCTCATCAGAACAGCGGGAAGCATCC A K M P P E S D V M T P P K V N V S T N G G P L S H Q N S G K H P	4900
4901	ACTGGGGAGCGTTCCCTCAGCAACAACAGCACAAAGCCCCGTGGGAAAGACAGACGCCAGTCTCTTTAGCTCAGGCTCCGGCAGTTGTGGGAAGTTCTAC L G S V P S A T T A Q S P V G K T D A S L F S S G S G S C G K F Y	5000
5001	AGTCCTCTCCCAATGACCAGCTGTTAAAAACACTGACTGAAAAGAACAGACAGTGGTTTAGCCTTTTGCCAAAAACACCTTGCGATGACACGTCCCTGA S P L P N D Q L L K T L T E K N R Q W F S L L P K T P C D D T S L T	5100
5101	CCCACGCTGACCTGTCCACCACTCTGGTGACTCCTCAGTCTCAGCCACCATCTAAGTCACCTTCGCCGGCCCCAGCTGCACTGCTCGGACCCTCTTCTGT H A D L S T T L V T P Q S Q P P S K S P S P A P A A L L G P S S V	5200
5201	TCAGAGCCCCCTGGACTGAACCCGTTTGCTCTGTGCGCTCTTCAGGTGAAAGGTGGTGTGTCTATGATGGGACTTCAGTTCTGTGGGTGGCCTGCTGGA Q S P P G L N P F A L S P L Q V K G G V S M M G L Q F C G W P A G	5300
5301	GTGCTCGCTTCCAACGTCCCATTTACATCACCTTTGCGCTGCTCTCGGCTCAGGGTTGGGACTGCCAGAAGGAAATGGCAGTAGTTTCATTCTTGACTTCCA V L A S N V P F T S P L P A L G S G L G L P E G N G S S S F L T S S	5400

5401	GTGTTGCTTCAAGTAAAGTGACTCTCCAGTCCCTCCAGCTGAAAGACCTCCAGCGCTCAGCCAGTGGCTGTTGAAGTAGCAAAGCCAGTAGATTTTCC V A S S K S D S P V P P A E R P S S A Q P V A V E V A K P V D F P	5500
5501	TAGTCCCAAACCTATCCCAGAAGAAATGCAGTTCGGATGGTGGAGAATCATTGACCCCGAGGACCTGAAGACATTGCTCAAAGTGCTGCACCTCCGCGGG S P K P I P E E M Q F G W W R I I D P E D L K T L L K V L H L R G	5600
5601	ATCCGGGAAAAGGCATTGCAAAAGCAAATTCAGAAACACTTGGATTACATTACCCAAGCCTGTGTCAGGAACAAGGACGTCGCTATTATTGAATTAAACG I R E K A L Q K Q I Q K H L D Y I T Q A C V R N K D V A I I E L N E	5700
5701	AAAATGAAGACAATCAGGTAACCCGCGATCTAGTGGAGAACTGGTCGGTCGAAAAGCAAGCAATGGAGCTGGACCTGAGCATACTCCAGCAGGTGGAGGA N E D N Q V T R D L V E N W S V E K Q A M E L D L S I L Q Q V E D	5800
5801	CTTGGAACGGAGGGTTGCATCAGCGAGTTTGCAAGTGAAGGGTTGGATGTGTCCAGAGCCTGCATCGGAAAGGGAGGACTTGGTATATTTTGAACATAAG L E R R V A S A S L Q V K G W M C P E P A S E R E D L V Y F E H K	5900
5901	TCACTTACTAAGTTGTTCAAGGAGCATGATGGAGAACTTACTGGTGACGAAGAAAACAGTGGCATGCACTGGCTCGGAAGAGTGACAACCCCTAGATA S L T K L F K E H D G E L T G D E E N S A H A L A R K S D N P L D I	6000
6001	TAGCTGTAACCAAGGCTGGCTGATTGGAGCGGAACATTGAGAGAAGGTATCTGAAGAGCCCTTAAGTACCACCATTGAGATCAAACCTGGATAATGTGGG A V T R L A D L E R N I E R R Y L K S P L S T T I Q I K L D N V G	6100
6101	CACAGTTACTGTCCCTGCTCCTGCACCATCCATTAGTGGTGATGGTGACGGAATTGAAGAGGATATTGCTCCAGGGCTCAGGGTCTGGAGAAGGGCCTTA T V T V P A P A P S I S G D G D G I E E D I A P G L R V W R R A L	6200
6201	GCAGAAGCTCGCAGTGCTGCGCAGGTAGCCCTGTGCATCCAGCAGTTACAGAGATCAATAGCATGGGAAAAATCAATTATGAAAGTTTACTGCCAGATCT A E A R S A A Q V A L C I Q Q L Q R S I A W E K S I M K V Y C Q I C	6300
6301	GTCGGAAGGGGACAAATGAAGAGCTGCTTCTACTCTGCGACGGCTGTGACAAAGGCTGCCACACCTACTGCCACAGGCCCAAGATTACAACCTATCCCCGA R K G D N E E L L L L C D G C D K G C H T Y C H R P K I T T I P D	6400
6401	TGGGGACTGGTTTTGCCCAGCTTGCATCTCTAAAGCAAGCGGCCAGAGTATAAAAATTAAAAAAATTCATGTCAAAGGAAAAAGACAAATGATTCCAAG G D W F C P A C I S K A S G Q S I K I K K I H V K G K K T N D S K	6500
6501	AAAACCAAGAAAGGAAACGTCGCCGGGGACACAGAGGATGAAGACTCGGCTTCCACAAGCAGCTCCCTCAAGAGAGGAAGCAAAGAGCTCAAAAAGAGGA K T K K G N V A G D T E D E D S A S T S S S L K R G S K E L K K R K	6600
6601	AAATGGAGGAAACCACTTCCCTCAACCTGTCAAAGCAGAAAGCACCACTCCATTAAAGAAACCCAAAAAGATGAGTCCAGGGACCTGGCTCTCTGCAG M E E T T S L N L S K A E S T T S I K K P K K D E S R D L A L C S	6700

6701	TATGATCCTGACTGAAATGGAGACACATGAGGACTCCTGGCCTTTCCTGCTTCCTGTGAACTTGAACTCGTACCCGGCTATAAGAAGGTTATCAAGAAA	6800
	M I L T E M E T H E D S W P F L L P V N L K L V P G Y K K V I K K	
6801	CCTATGGATTTTCTACAATTAGAGAAAACTAAATAATGGACAGTATCCCAACTTTGAGACCTTGCTCTAGACGTCAGGCTTGTTTTGACAACTGTG	6900
	P M D F S T I R E K L N N G Q Y P N F E T F A L D V R L V F D N C E	
6901	AAACATTTAATGAAGATGACTCGGACATAGGCAGAGCTGGCCACAGTATGCGAAAATACTTTGAAAAAAGTGGACAGATACCTTCAAAGTGAGCTGAAG	7000
	T F N E D D S D I G R A G H S M R K Y F E K K W T D T F K V S *	
7001	TTGTAATCTCCTTTTCTCCTTTTTCCTTTTAAACAAGGACAAACAAGACCAGCGGTGTTGAACTGCATTACACAGATGT	7081

### 3.2.4 Structure and splice variants of the mPBM1 cDNA

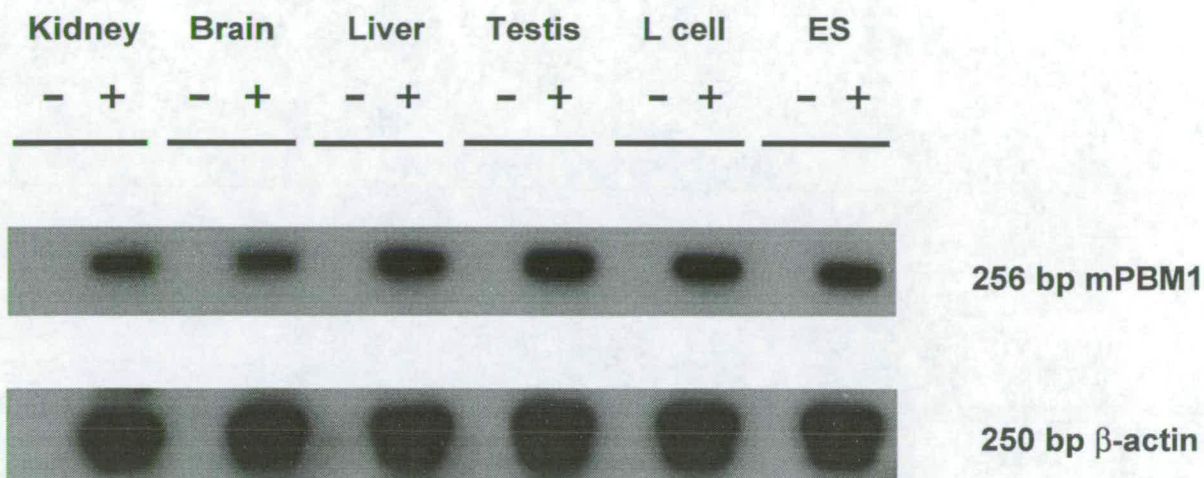
The cDNA sequence derived by alignment of the  $\lambda$ ZK12, p519459 and 363697 clones is 7081 nucleotides in length and contains a long open reading frame (see Figure 3.5 for mPBM1 cDNA and amino acid sequences). Kozak has defined the optimal initiator methionine consensus sequence for translation, in vertebrate mRNAs as GCCACCAUGG (Kozak, 1986, 1987) and it has been demonstrated that two positions are highly conserved : a purine (normally A) at position -3 and a G at +4 (where A of AUG is +1). The ATG at position 411 of the mPBM1 cDNA was chosen as the initiator methionine codon for the conceptual translation for a number of reasons. Firstly, the putative initiator methionine codon at position 411 of the mPBM1 cDNA sequence conforms to the R<sup>-3</sup> (R=purine) and G<sup>+4</sup> consensus for translational initiation described above. Secondly, numerous upstream terminator codon are evident within the 5'UTR of the mPBM1 cDNA and the ATG at position 411 represents the first in-frame methionine codon. However, in order to determine whether a cDNA sequence is complete, a favourable context for initiation around a potential ATG start codon and the presence of an upstream, in-frame stop codon are insufficient grounds for the assignment of an initiator methionine. Kozak (1996) has pointed out the presence of these two positive indicators cannot represent the sole criteria for assessing the completeness of a new cDNA sequence and the absence of certain negative indicators is also required. For example, an upstream out of frame start codon that generates a reading frame that overlaps the major open reading frame, strongly suggests that an error in sequencing of the cDNA may have occurred. Within



the mPBM1 5'UTR, no open reading frames are evident upstream of the putative methionine at position 411 and furthermore the out of frame ATG codons are not in favourable contexts for translational initiation. Whilst these indicators suggest the ATG codon at position 411 is the initiator methionine codon, to unambiguously map the 5' terminus of the mPBM1 message primer extension analysis would need to be carried out.

RT-PCR analysis was carried out on cDNA generated from total RNA from a range of mouse tissues and cell lines to ascertain the expression pattern of the mPBM1 transcript. Figure 3.6 shows the results of this expression analysis of mPBM1 following Southern blotting and hybridisation with an internal labeled oligonucleotide specific to the mPBM1 transcript. mRNA transcripts of the correct size that hybridise to an internal oligomer were detectable in all tissues and cell lines analysed but were not present in preparations where no reverse transcriptase was added. These were confirmed to be mPBM1 transcripts by direct sequencing of DNA fragments. These results taken together with the widespread origin of ESTs derived from the mPBM1 cDNA suggest a ubiquitous pattern of expression for mPBM1.

During the isolation, restriction mapping and sequencing of the mPBM1 cDNA clones, a series of alternatively spliced isoforms of the mPBM1 transcript were identified. Figure 3.5 shows the positions of the splice variants within the mPBM1 cDNA and the amino acids encoded by these splicing events. The predicted structure



**Figure 3.6** : Analysis of mPBM1 expression in murine tissues and cell lines

Southern blot hybridisation following RT-PCR analysis from cDNA prepared from kidney, brain, liver, testis, L929 fibroblasts and undifferentiated embryonic stem (ES) cells. Results are presented as an autoradiograph following hybridisation with an internal labeled mPBM1-specific oligonucleotide. A  $\beta$ -actin positive control was performed on the same cDNA samples and hybridised to an internal labeled  $\beta$ -actin-specific oligonucleotide.

zZKEXP/F	: 5'-CGTTAACGAGCTGGCATGCAGC-3'
mZKEXP/R	: 5'-ATCATCGTCATAATCACTGTCC-3'
mZKEXP-INT	: 5'-AGCTTAGGATAATTCATGCT-3'
MUSBETA-ACTIN/F	: 5'-ACACCTTCTACAATGAGCTGCG-3'
MUSBETA-ACTIN/R	: 5'-AGAGCATAGCCCTCGTAGATGG-3'
MUSBETA-ACTIN-INT	: 5'-AGCCATGTACGTAGCCATCC-3'

- and + denotes no addition or addition of reverse transcriptase during preparation of the cDNA.

of the proteins encoded by these alternative splicing events is diagrammed in Figure 3.7. Three major RNA isoforms of the mPBM1 transcript were detected and are termed mPBM1a, mPBM1b and mPBM1c. mPBM1a splices out a 6bp region at the extreme 5' end of the mPBM1 cDNA (at nucleotide positions 792-798) and interestingly removes a 285bp region (positions 2360-2645) immediately 5' of the MBD. The mPBM1b isoform splices out the 6bp and 285bp sequences as in the mPBM1a transcript but also removes a 27bp region 3' of the MBD (nucleotide positions 3358-3385). Finally, the mPBM1c transcript alternatively splices a 66bp sequence adjacent to the MBD (cDNA nucleotide 3246-3312) (see Figure 3.5).

It was of interest to determine whether spliced and unspliced mRNAs corresponding to the 285bp event adjacent to the MBD of mPBM1 are detectable in murine tissues or cultured cells given the proximity of the alternative splicing event to the MBD. RT-PCR analysis using primers flanking the predicted site of the 285bp splicing event were used to amplify transcripts from a range of somatic and mouse cell line cDNAs (see Figure 3.7C). Southern blot hybridisation with an internal labelled oligomer specific to the mPBM1 transcript, the amplification of products of the predicted size and the direct sequencing of these fragments confirmed the 285bp alternative splicing events in the mPBM1 isoforms. Furthermore, the identification of both spliced and unspliced versions of the mPBM1 transcripts from independent cDNA clones strongly suggests these represent genuine alternative splicing events and are not simply artefacts generated during construction of the cDNA library from

### **Figure 3.7 : Structure and expression of alternative splicing events in mPBM1**

#### **A. Alignment of the mPBM1a, mPBM1b and mPBM1c splice variants**

Alternative splicing events are indicated by diagonal lines below the protein isoforms. The scale bar at the top of the diagram corresponds to amino acid positions numbered from the N-terminus.

#### **B. Alternative splicing events within the mPBM1 cDNA**

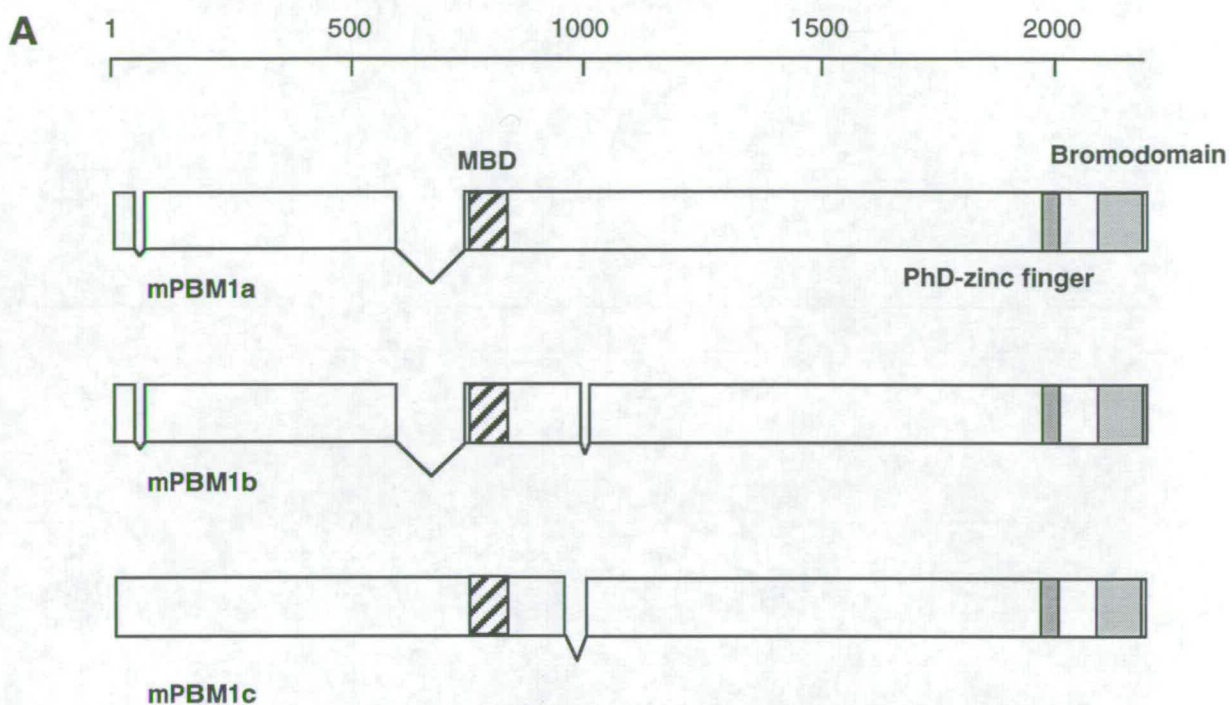
mPBM1a splices out a 6bp region at the extreme 5' end of the mPBM1 cDNA (at cDNA nucleotide positions 792-798) and interestingly removes a 285bp region (positions 2360-2645) immediately 5' of the MBD. The mPBM1b isoform contains the 6bp and 285bp alternative splicing events seen in the mPBM1a transcript and also removes a 27bp region 3' of the MBD (at positions 3358-3385). Finally, the mPBM1c transcript alternatively splices out a 66bp sequence adjacent to the MBD (positions 3246-3312).

#### **C. RT-PCR analysis of the mPBM1 285bp splice isoform in murine tissues and cell lines**

Southern blot hybridisation with a probe specific to mPBM1 following RT-PCR analysis from cDNA prepared from kidney, brain, liver, testis, L929 fibroblasts and undifferentiated embryonic stem (ES) cells using mPBM1 primers flanking the predicted 285bp splice site (mPBM1 +285/F and mPBM1 +285/R). Results are presented as an autoradiograph following hybridisation with an internal <sup>32</sup>P-labeled mPBM1-specific oligonucleotide (mPBM1-INT). Spliced and unspliced isoforms of mPBM1 are detectable in all cases.

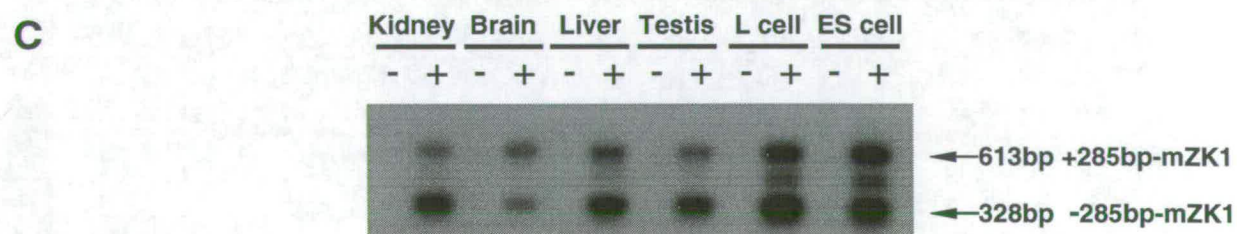
- and + denotes no addition or addition of reverse transcriptase during preparation of the cDNA.

mPBM1+285/F	5'-TGAAGAGGATGATGAAGATGATGAG-3'
mPBM1+285/R	5'-GGATTCCTCTCTTGCTCTGG-3'
mPBM1-INT	5'-GATGTCATTCCTCGTATCAGG-3'



**B**

Splice Isoform	+6bp	+285bp	+66bp	+27bp
mPBM1a	—	—	+	+
mPBM1b	—	—	+	—
mPBM1c	+	+	—	+



which they were isolated. In order to unambiguously demonstrate that these mPBM1 isoforms were attributable to alternative splicing events, genomic clones need to be isolated and the splice donor and splice acceptor consensus sequences identified.

The functional significance of these alternative splice isoforms is unknown but interestingly three of these splicing events are located either immediately 5' or 3' of the MBD of mPBM1. Bickmore et al., (1992) have demonstrated that the DNA-binding properties of the WT1 transcription factor are modulated by alternative splicing events within a zinc finger region of this protein. The DNA binding domains encoded by these different RNA species of WT1 were found to possess distinct affinities for DNA sequences in vitro and also conceivably for downstream physiologic targets in vivo. Similiar findings have been reported for the DNA-binding proteins E12 and E47, products of alternative splicing events of the *E2A* gene that possess distinct affinities for sequence elements in vitro (Sun and Baltimore, 1991).

### **3.2.5 Identification of a family of mammalian PBM proteins**

As described in section 3.2.3 a human EST (363697) was successfully employed as a heterologous probe to isolate cDNA clones spanning the 3' end of the mPBM1 cDNA. This human cDNA clone was employed as a cross-species probe because it encoded a PHD-zinc finger and bromodomain motif highly homologous to the *C. elegans* PBM-1 cDNA sequence ( $P < 10^{-24}$ ). It is important to note that this human cDNA clone contained significant sequence identity outside these conserved regions

to the *C. elegans* PBM-1 sequence because as the PHD-zinc finger and bromodomain motifs are present in many other eukaryotic proteins the presence of a bipartite PHD-zinc finger/bromodomain motif *per se* in a EST would not be sufficient grounds for justifying relatedness to the nematode protein. Furthermore, a cluster of human ESTs that contig with the 363697 human cDNA clone were identified by searching dbEST, XREF and Unigene databases (<http://www.ncbi.nlm.nih.gov/Schuler/Unigene>) and a partial cDNA sequence was assembled and is referred to as hPBM1 for human PHD-zinc finger, Bromodomain and MBD1.

A human cDNA sequence (KIAA0314), that is importantly distinct from the hPBM1 cDNA described above, has been deposited in GenBank (Accession No. AB002312) that potentially encoded an additional PHD-zinc finger/bromodomain-containing protein with a highly significant sequence identity to the *C. elegans* PBM-1 protein ( $P < 10^{-37}$ ). This cDNA sequence was as a result termed hPBM2 to distinguish it from hPBM1 described above.

To determine whether a hPBM2 homolog existed in the mouse, primers specific to the hPBM2 cDNA were employed to amplify a 780bp fragment corresponding to the PHD zinc finger/bromodomain region from HeLa cDNA :

KIAA0314F: 5'-TTCGTGTCTGGCGCCAGACCC-3'

KIAA0314R: 5'-TCCCTGATAAACTCCTCCC-3'

This 0.7kb fragment was used to screen the mouse pituitary  $\lambda$ ZAPII cDNA library. From this screen seven pure phage plaques were identified, in vivo excised, restriction digested and the fragments sized on agarose gels. These cDNA clones were sequenced from each end by sequence cycling using fluorescent dye terminators (Perkin-Elmer). The largest of these clones ( $\lambda$ 105) contained a 2.5kb cDNA insert and was sequenced entirely. Conceptual translation of this partial  $\lambda$ 105 cDNA and BLAST and FASTA searches against peptide sequence databases revealed a highly significant protein sequence identity to the predicted protein product from the hPBM2 (KIAA0314) partial cDNA sequence and the *C. elegans* PBM-1 protein ( $P < 10^{-41}$ ). This  $\lambda$ 105 murine cDNA is termed mPBM2 for murine PHD-zinc finger, Bromodomain and MBD2 to distinguish it from the mPBM1 cDNA.

An alignment of the predicted protein products from the complete cDNAs from *C. elegans* PBM-1 and mPBM1 and from the partial cDNA sequences for hPBM1, mPBM2 and hPBM2 cDNAs is shown in Figure 3.8. The PHD-zinc finger motifs of *C. elegans* PBM-1, mPBM1, mPBM2, hPBM1 and hPBM2 and related PHD-zinc finger proteins is presented in Figure 3.9. The bromodomains of *C. elegans* PBM-1, mPBM1, mPBM2, hPBM1 and hPBM2 and the related bromodomains of yeast GCN5 and the human PCAF proteins are presented in Figure 3.10. These findings suggest the possible existence of a gene family in vertebrates, with two human and mouse genes encoding proteins highly homologous to the *C. elegans* PBM-1 protein, the product of a single gene extant in the sequenced genome of this invertebrate.





**Figure 3.9 :** Sequence comparison of the PHD-zinc finger motifs (C4HC3) of the PBM proteins with PHD-zinc finger containing proteins

Protein sequence homology was identified by BLASTp database searches. The shaded residues represent amino acid sequence identity or presence of similar amino acids in greater than 50% of cases. The PHD (plant homeodomain) zinc finger, also termed TTC (trithorax consensus) and LAP (leukaemia-associated-protein) is present in known or suspected chromatin-associated proteins and is thought to represent a protein-protein interaction motif (Aasland et al., 1995). The PHD-zinc finger proteins in the alignment are grouped according to the degree of sequence conservation of the non-zinc co-ordinating residues.

hWTSF is a putative transcription factor deleted in Williams syndrome; hMOZ is the monocytic leukaemia zinc finger protein; hSMCX escapes X-chromosome inactivation and hSMCY is the pseudoautosomal homolog; RBBP2 is a retinoblastoma-binding protein; AIRE is a putative transcriptional regulator mutated in the autoimmune disorder APECED; Mi-2 is a component of the NuRD nucleosome remodeling and histone deacetylase complex (Mi-2a and Mi2b are the two PHD fingers of the protein); KAP1(TIF1 $\beta$ ) is a transcriptional co-repressor which like TIF1 $\alpha$  can interact with mammalian homologs of the *Drosophila* heterochromatin protein-1.

PBM-1	C	Q	I	C	K	.	S	M	D	G	D	E	.	M	L	V	C	D	G	C	E	S	G	C	H	M	E	C	F	R	P	R	M	T	K	V	P	E	G	D	W	F	C	Q	R	C
mPBM1	C	Q	I	C	R	.	K	G	D	N	E	E	L	L	L	L	C	D	G	C	D	K	G	C	H	T	Y	C	H	R	P	K	I	T	T	I	P	D	G	D	W	F	C	P	A	C
hPBM1	C	Q	I	C	R	.	K	G	D	N	E	E	L	L	L	L	C	D	G	C	D	K	G	C	H	T	Y	C	H	R	P	K	I	T	T	T	P	D	G	D	W	F	C	P	A	C
mPBM2	C	L	V	C	R	.	K	G	D	N	D	E	F	L	L	L	C	D	G	C	D	R	G	C	H	I	Y	C	H	R	P	K	M	E	A	V	P	E	G	D	W	F	C	A	V	C
hPBM2	C	L	V	C	R	.	K	G	D	N	D	E	F	L	L	L	C	D	G	C	D	R	G	C	H	I	Y	C	H	R	P	K	M	E	A	V	P	E	G	D	W	F	C	T	V	C
hWTSF	C	K	V	C	R	.	K	K	G	E	D	D	K	L	I	L	C	D	E	C	N	K	A	F	H	L	F	C	L	R	P	A	L	Y	E	V	P	D	G	E	W	Q	C	P	A	C
hMOZ	C	S	S	C	R	D	Q	G	K	N	A	D	N	M	L	F	C	D	S	C	D	R	G	F	H	M	E	C	C	D	P	P	L	T	R	M	P	K	G	M	W	I	C	Q	I	C
hSMCX	C	Q	V	C	S	.	R	G	D	E	D	D	K	L	L	F	C	D	G	C	D	D	N	Y	H	I	F	C	L	L	P	P	L	P	E	I	P	R	G	I	W	R	C	P	K	C
hSMCY	C	Q	V	C	S	.	R	G	D	E	D	N	K	L	L	F	C	D	G	C	D	D	N	Y	H	I	F	C	L	L	P	P	L	P	E	I	P	R	G	I	W	R	C	P	K	C
hRBBP2	C	M	F	C	G	.	R	G	N	N	E	D	K	L	L	L	C	D	G	C	D	D	S	Y	H	T	F	C	L	I	P	P	L	P	D	V	P	K	G	D	W	R	C	P	K	C
hAIRE	C	A	V	C	R	D	G	G	E	.	.	.	.	L	I	C	C	D	G	C	P	R	A	F	H	L	A	C	L	S	P	P	L	R	E	I	P	S	G	T	W	R	C	S	S	C
hMi2 $\alpha$	C	E	V	C	Q	Q	G	G	E	.	.	.	.	I	I	L	C	D	T	C	P	R	A	Y	H	M	V	C	L	D	P	D	M	E	K	A	P	E	G	K	W	S	C	P	H	C
hMi2 $\beta$	C	R	V	C	K	D	G	G	E	.	.	.	.	L	L	C	C	D	T	C	P	S	S	Y	H	I	H	C	L	N	P	P	L	P	E	I	P	N	G	E	W	L	C	P	R	C
hKAP1	C	R	V	C	Q	K	P	G	D	.	.	.	.	L	V	M	C	N	Q	C	E	F	C	F	H	L	D	C	H	L	P	A	L	Q	D	V	P	G	E	E	W	S	C	S	L	C
hTIF1 $\alpha$	C	A	V	C	Q	N	G	G	E	.	.	.	.	L	L	C	C	E	K	C	P	K	V	F	H	L	S	C	H	V	P	T	L	T	N	F	P	S	G	E	W	I	C	T	F	C



PBM-1	L	M	L	D	E	L	V	V	Q	A	N	A	L	P	F	L	E	P	V	N	P	K	L	V	P	G	Y	K	M	I	I	S	K	P	M	D	L	K	T	I	R	Q	K	N	E	K	L	I	Y	E	T	P	E	D	F	A	E	D	I	E
mPBM1	M	I	L	T	E	M	E	T	H	E	D	S	W	P	F	L	L	P	V	N	L	K	L	V	P	G	Y	K	K	V	I	K	K	P	M	D	F	S	T	I	R	E	K	L	N	N	G	Q	Y	P	N	F	E	T	F	A	L	D	V	R
hPBM1	M	I	L	T	E	M	E	T	H	E	D	A	W	P	F	L	L	P	V	N	L	K	L	V	P	G	Y	K	K	V	I	K	K	P	M	D	F	S	T	I	R	E	K	L	S	S	G	Q	Y	P	N	L	E	T	F	A	L	D	V	R
mPBM2	I	I	L	M	E	M	E	S	H	D	A	A	W	P	F	L	E	P	V	N	P	R	L	V	S	G	Y	R	R	V	I	K	N	P	M	D	F	S	T	M	R	E	R	L	L	R	G	G	Y	T	S	S	E	E	F	A	A	D	A	L
hPBM2	I	I	L	M	E	M	E	S	H	D	A	A	W	P	F	L	E	P	V	N	P	R	L	V	S	G	Y	R	R	I	I	K	N	P	M	D	F	S	T	M	R	E	R	L	L	R	G	G	Y	T	S	S	E	E	F	A	A	D	A	L
ScGCN5	N	I	L	T	E	L	Q	N	H	A	A	A	W	P	F	L	Q	P	V	N	K	E	E	V	P	D	Y	Y	D	F	I	K	E	P	M	D	L	S	T	M	E	I	K	L	E	S	N	K	Y	Q	K	M	E	D	F	I	Y	D	A	R
hPCAF	S	I	L	Q	Q	V	K	S	H	Q	S	A	W	P	F	M	E	P	V	K	R	T	E	A	P	G	Y	Y	E	V	I	R	S	P	M	D	L	K	T	M	S	E	R	L	K	N	R	Y	Y	V	S	K	K	L	F	M	A	D	L	Q

PBM-1	L	M	F	A	N	C	R	Q	F	N	I	D	H	S	E	I	G	R	A	G	I	S	L	H	K	F	F	Q	K	R	W	K	Q	L	K	Y
mPBM1	L	V	F	D	N	C	E	T	F	N	E	D	D	S	D	I	G	R	A	G	H	S	M	R	K	Y	F	E	K	.	W	T	D	T	F	
hPBM1	L	V	F	D	N	C	E	T	F	N	E	D	D	S	D	I	G	R	A	G	H	N	I	E	E	V	F	W	K	K	V	G	P	D	T	F
mPBM2	L	V	F	D	N	C	Q	T	F	N	E	D	D	S	E	V	G	K	A	G	H	V	M	R	R	F	F	E	S	R	.	W	E	E	F	Y
hPBM2	L	V	F	D	N	C	Q	T	F	N	E	D	D	S	E	V	G	K	A	G	H	I	M	R	R	F	F	E	S	R	.	W	E	E	F	Y
ScGCN5	L	V	F	N	N	C	R	M	Y	N	G	E	N	T	S	Y	Y	K	Y	A	N	R	L	E	K	F	F	N	N	K	V	K	E	I	P	E
hPCAF	R	V	F	T	N	C	K	E	Y	N	A	P	E	S	E	Y	Y	K	C	A	N	I	L	E	K	F	F	F	S	K	I	K	E	A	G	L

**Figure 3.10** : Sequence comparison of the PBM bromodomains

Homology alignment of the bromodomains from the *C. elegans* pbm-1 and mammalian proteins with the yeast GCN5 and human PCAF sequences. Shaded residues represent amino acid sequence identities or conservative substitutions in greater than 50% of cases.

**Figure 3.11** Sequence alignment of the mPBM1 and hPBM1 proteins

Shaded residues represent amino acid sequence identities or conservative substitutions events. The mPBM1 protein is derived from a complete cDNA whereas the hPBM1 sequence was obtained from a partial cDNA clone. The 35 amino acid insertion event between residues 1879 and 1914 in the mPBM1 protein may correspond to an unidentified splice isoform. Interestingly, it is immediately adjacent to the PHD-zinc finger region (amino acids 1961-2004) and the bromodomain motifs (2098-2191).

The PBM1 proteins are highly homologous to one another within their PHD-zinc finger and bromodomain regions and also crucially contain high sequence identity outside these regions not present in the related PHD-zinc finger, bromodomain and PHD-zinc finger/bromodomain-containing proteins. Database searches with mPBM1 and hPBM1 revealed highly significant sequence identity to the *C. elegans* PBM-1 protein and slightly more limited sequence identity to the mPBM2 and hPBM2 proteins.



mPBM1	1712	RIIDPEDLK	TLLKVLHLRGIREKALQKQIQKHLDYITQAC	VRNKDVAIE
hPBM1	1	RIIDPEDLK	ALLKVLHLRGIREKALQKQIQKHLDYITQAC	LKNKDVAIE
	1762	LNENE	DNQVTRDLVENWSVEKQAMELDLSILQQVEDLERRVASASLQVKG	
	50	LNENE	ENQVTRDIVENWSVEEQAMEMDLSVLQQVEDLERRVASASLQVKG	
	1812	WMCPEPASEREDLVYFEHKS	LTKL	FKEHDGELTGDEENSAHALARKSDNP
	100	WMCPEPASEREDLVYFEHKS	FTKL	CKEHDGEFTGEDESSAHALERKSDNP
	1862	LDIAVTRLADLERNIERRY	LYLKSPLSTTIQIKLDNVGTVTV	PAPAPSGD
	150	LDIAVTRLADLERNIERR	.....	
	1912	GDG	IEEDIAPGLRVWRRALA	EARSAAQVALCIQQLQRSIAWEKSIMKVYC
	200	...	IEEDIAPGLRVWRRALS	EARSAAQVALCIQQLQKSIAWEKSIMKVYC
	1962	QICRKGDNEELLLL	CDGCDKGCHTYCHRPKIT	TIPDGDWFCPACISKASG
	250	QICRKGDNEELLLL	CDGCDKGCHTYCHRPKIT	TIPDGDWFCPACIAKASG
	2012	QSIKIKKI	HVKGKKTND	SKKTKKGNVAGDTEDEDSASTSSSLKRGSKELK
	300	QTLKIKKL	HVKGKKTNE	SKKGKKVTLTGDTEDSDASTSSSLKRGNKDLK
	2062	KRKMEET	TSLNLSKAEST	TTSIKKPKKDES
	350	KRKMEENT	SINLSKQES	FTSVKKPKRDDSKDLALCSMILTEMETHEDSWP
	2112	FLLPVNLKLVP	GYKKVIKKPMD	FSTIREKLNN
	400	FLLPVNLKLVP	GYKKVIKKPMD	FSTIREKLSS
				GOYPNLET
				FALDVRLVFD
	2162	NCETFNEDDSDIGRAGH	S	MRKYFEKKWTDTFKVS
	450	NCETFNEDDSDIGRAGH	N	MRKYFEKKWTDTFKVS

**Figure 3.12** Sequence alignment of the mPBM2 and hPBM2 proteins

Shaded residues represent amino acid sequence identities or conservative substitution events. The mPBM2 and hPBM2 protein sequences are derived from partial cDNA clones. They are highly homologous within their PHD-zinc finger and bromodomain regions but also importantly contain sequence identity outwith these regions that are not conserved in related proteins.



mPBM2 1 SLGQSQHDL SQSAFLSWLSQTQSHN SLLSSSVLTPDSSPGKLD SAPSQS LEEPEPDEAQ SCFPGPQGPFWN  
hPBM2 1 SLGQSQHDL SQSAFLSWLSQTQSHS SLLSSSVLTPDSSPGKLD PAPSQPPEEPEPDEAESSPDQPALWFN

71 FSAQIPCD AAPTPPPAVSEDQPTPS LQL LASSKPMNTP GAANPCSPVQL SSTHLPGGT PKRLSGDSEEMS  
71 ISAQMPCNAAPTPPPAVSEDQPTPS PQQLASSKPMNRPSAANPCSPVQF SSTPLAGLAPKRRAGDPGEMP

141 QSPTGLGQPKRRGRPPSKFFKQVEQH YLTQLTAQPI PPEMCSGWWIRDPE TLDVLLKALHPRGIREKAL  
141 QSPTGLGQPKRRGRPPSKFFKQMEQRYLTQLTAQPV PPEMCSGWWIRDPEMLDAMLKALHPRGIREKAL

211 HKHLSKHKDFLQEVCLQPLTDPIFEPNELPALEEGVMSWSPKEKTYETDLAVLQWVEELEQRVVLSDLQI  
211 HKHLNKHDFLQEVCLRPSADPIFEP RQLPAFQEGIMS WSPKEKTYETDLAVLQWVEELEQRVIMSDLQI

281 RGWTCPTPDSTREDLTYCEHLPDSPEDIPWRGRGREGTVPQRQNNNPLDLAVMRLAVLEQNVERRYLREP  
281 RGWTCPSPDSTREDLAYCEHLSDSQEDITWRGRGREGLAPQRKTTNPLDLAVMRLAALEQNVERRYLREP

351 LWAAHEVVVEKALLSTPNGAPDGTSTEISYEITPRVRVWRQTLERCRSAAQVCLCMGQLERSIAWEKSVN  
351 LWPTHEVVLEKALLSTPNGAPEGTTTEISYEITPRI RVWRQTLERCRSAAQVCLCLGQLERSIAWEKSVN

421 KVTCLVCRKGDNDEFLLLCDGCDRGCHIYCHRPKMEAVPEGDWFC AVCLSQQVEEEYTQRPGF PKRGQKR  
421 KVTCLVCRKGDNDEFLLLCDGCDRGCHIYCHRPKMEAVPEGDWFC TVCLAQQVEGEFTQKPGF PKRGQKR

491 KSSFPLTFPEGDS.RRRMLSRSRDSPAV.PRYPEDGLSPPKRRRHSMRSHHSDLTFCEIILMEMESHDA  
491 KSGYSLNFSEGDGRRRRVLLRGRES PAAGPRYSEEGLSPPSKRRRLSMRNHHSDLTFCEIILMEMESHDA

559 WPFLEPVNPRLVSGYRRV IKNPMDFSTM RERLLRGGYTSSEEF AADALLVFDNCQTFNEDDSEVGKAGHV  
561 WPFLEPVNPRLVSGYRR IKNPMDFSTM RERLLRGGYTSSEEF AADALLVFDNCQTFNEDDSEVGKAGHI

629 MRRFFESRWEEFYQGKQANL  
631 MRRFFESRWEEFYQGKQANL



### **3.2.6 Construction of bacterial expression vectors and the purification of PBM-1, mPBM1 and MBD4 MBDs**

The N-terminal MBD of the *C. elegans* PBM-1 and murine PBM1 proteins are homologous to the MBDs of known methyl-CpG binding proteins (see Figure 3.13). In order to determine whether these motifs were capable of binding specifically to methylated template, a set of expression vectors were constructed encoding the MBDs of the *C. elegans* PBM-1 and mouse PBM1 proteins. These vectors were expressed in *E. coli* as recombinant proteins, affinity purified and tested for their ability to associate with methylated and unmethylated probes in a series of in vitro binding assays. Nan et al., (1993, 1996) have previously demonstrated that an 80 amino acid N-terminal region of MeCP2, the MBD is required for the specific association with methylated sites in vitro and in vivo. Furthermore, this minimal domain expressed as a fusion protein can discriminate methylated and unmethylated templates in a directly analogous fashion to the full length protein.

In order to express the MBD of *C. elegans* PBM-1 and mouse PBM1 as recombinant proteins, the cDNAs encoding these motifs were subcloned into a bacterial expression vector, pGEH. This expression construct contains the cDNA encoding glutathione-S-transferase (GST) upstream of a multiple cloning site and a downstream sequence coding for six histidine residues, allowing protein purification by either glutathione-Sephadex beads or by nickel ion affinity chromatography. A 412bp cDNA product encoding the MBD of *C. elegans* PBM-1 and a 543bp cDNA product encoding the MBD of the mPBM1 protein were amplified by PCR so as to

incorporate 5' BamHI and 3' MfeI sites for subcloning to a BamHI/EcoRI linearised pGEH expression vector to generate the expression constructs, pGEH-PBM-1-MBD and pGEH-mPBM1-MBD respectively :

BamHI-PBM-1 MBD/F : 5'-GCACTCGGATCCAAAGTGCTTGCAGCTCACC-3'  
MfeI-PBM-1 MBD/R : 5'-CGAATTCAATTGCATTTAATCTTGTAAGTTCC-3'

BamHI-mPBM1 MBD/F : 5'-AGTCTCGGATCCCACTCAGCACCTCACCTC-3'  
MfeI-mPBM1 MBD/R : 5'-TTGTGCAATTGGTCTGGCTATTTCTTGAGC-3'

These expression constructs were sequenced to ensure that the PBM-1 and mPBM1 cDNAs were subcloned in-frame and that no mutations were present. These vectors were electrotransformed into *E. coli* BL21 (DE3) pLysS, grown in LB media until OD<sub>600</sub> was 0.4 and protein expression was induced with 0.4mM isopropylthio- $\beta$ -D-galactosidase (IPTG) at for 3 hrs. Upon induction, these pGEH expression constructs would generate a recombinant fusion protein with a glutathione S-transferase (GST) tagged N-termini fused in-frame to the MBD and a C-terminal tag of six histidine residues. The induced bacterial pellets were resuspended in a high salt buffer (300mM NaCl) containing protease inhibitors and the cells were lysed by sonication. These extracts were then passed over a nickel agarose column (Qiagen) then washed extensively with a buffer containing Triton-X-100 (0.1%) and the histidine analog, imidazole (20mM) to minimise non-specific protein adsorption. The

### Figure 3.13 Sequence comparison of the MBD protein family

Amino acid sequence alignment of the MBD regions from *C. elegans* PBM-1, mouse PBM1 and a human EST (593380) highly homologous to mPBM1. The MBDs of mouse and human MeCP2, MBD 1, MBD2, MBD4 and MBD4 and plant and invertebrate MBD-containing proteins are included. The numbering scheme above the diagram is according to the MBD of MeCP2.

An NMR solution structure exists for the MBD of MeCP2 (Wakefield et al, 1999). Tyr123 and the adjacent charged residues Arg111, Asp121 and Arg133 nearby are thought to be involved in making contact with m<sup>5</sup>CpG. The high proportion of basically charged amino acids (R, K) flanking the central tyrosine residue at position 123 is consistent with the MBD motif associating with m<sup>5</sup>CpGs in a variety of sequence contexts where the flanking positively charged residues could potentially mediate electrostatic interactions with the negatively charged phosphodiester backbone of DNA. This Tyr 123 is absent in the *C. elegans* PBM-1, mPBM1 and the human EST 593380 strongly supporting the in vitro and in vivo findings that the MBDs from these proteins are non-functional.

II represents the location of a conserved intron within the MBD of the mammalian MBD1-4 and MeCP2 genes. This intron position is not conserved within the MBD of the nematode *pbm-1* gene which contains an intron situated at III location (see Table 3.1). Furthermore, during the mapping of the hPBM1 gene primers specific to the 593380 human EST which encodes an MBD highly homologous to the mPBM1 identified an intron at the position indicated by I (see Figure 5.2).

Species are indicated in the domain names as : Hs, human; Mm, mouse; Ce, *Caenorhabditis elegans*; Zm, maize; Os, rice; At, *Arabidopsis thaliana*; Am, snapdragon; Dr, Zebrafish; Bf, *Amphioxus*; Ad, cricket; Dv, *Drosophila melanogaster*; Dv, *Drosophila virilis*; Le, tomato.

III

	90	100	I	110	120	II	130	140	150	160
HsEST 593380	TDERELRIP.	LEYGWQRETRIRNFGGR.	LQGEVAYYAPC.	GKKLRQYPEVIRGM.....	QWCLLKEEDVIPRI	RAMEGRR				
MmPBM1	ADDQELRIP.	LDYGWQRETRVRNFGGR.	LPGEVAYYAPC.	GKKLRQCPDMVKG.....	QWCLLKEEDVIPRI	RAMDGRR				
CePBM-1	TNEAMLRIP.	LQLGWRROTCVRSIAS	AGVKGDSYFAPC.	GKKLSTYSEVVRYLTK..	NSIHYITRDNFLFNTKLVIGEF					
ZmMBD	KCASTTEIPA.	PSGWTKKLAPTRGG.	RGGRFEVIFVSP.	TGEEIKSKRQLTOYLKAHPGGA.	SSEFDWGTSDTPRRSAR					
OsMBD	KEVSVVEMPA.	PEGWTKKFTTPQ....	RGGRFEVIFVSP.	TGEEIKNKRQLSQYLKAHPGGA.	SSEFDWGTGDTTPRRSAR					
AtMBDa	DELVSIELPA.	PASWKLLFYPPKRA.	TPRKTEIVFVAP.	TGEEISSRKQLEQYLKAHPGNPV.	ISEFEWTTTGETTPRRSSR					
AmMBD	DDVSVVELPA.	PAAWKKLYLPKKG.	TPKKNEILFIAP.	TGEEISSRRQLEQYLKANPGNPA.	LSEFDWGTGETTPRRSAR					
AtMBDb	EEVSVVELPA.	PSSWKLLFYPPNVG.	SVKKTEVVFVAP.	TGEEISNRKQLEQYLKSHPGNPA.	IAEFDWTTTSGTPRRSAR					
MmMeCP2	DRGPMYDDPTLPE	GWTRKLKQKSGRSAGKYDVYLIP.	QGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGCGSPSRR							
HsMeCP2	DRGPMYDDPTLPE	GWTRKLKQKSGRSAGKYDVYLINP.	QGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGCGSPSRR							
MmMBD4	STTATEGHKPVPC	GWERVVKQRLSGKTAGKFDVYFISP.	QGLKFRSKRSLANYLLKNGETFLKPEDFNFTVLPRKGSINPG							
HsMBD4	ATAGTECRKSVPC	GWERVVKQRLFGKTAGKFDVYFISP.	QGLKFRSKSSLANYLLKNGETSLKPEDFDFTVLPRKGIKSR							
MmMBD1	MAESWQDCPALGP	GWKRRESFRKSGASFGRSDIYYQSP.	TGEKIRSKVELTRYL...GEA.CDLTLFDFR.QGTLCHPIP							
HsMBD1	MAEDWLDPCALGP	GWKRREVFRRKSGATCGRSDIYYQSP.	TGDRIRSKVELTRYL...GEA.CDLTLFDFR.QGILCYPAP							
MmMB2	ESGKRMDCPALPP	GWKKEEVIRKSGLSAGKSDVYYYFSP.	SGKKFRSKPQLARYL...GNAV.DLSSFDFFR.TGKMMPSKL							
HsMBD2	ESGKRMDCPALPP	GWKKEEVIRKSGLSAGKSDVYYYFSP.	SGKKFRSKPQLARYL...GNTV.DLSSFDFFR.TGKMMPSKL							
DrMBD2	MEKKRTDCPALPP	GWKKEEVIRKSGLSAGKSDVYYYFSP.	KGKKFRSKPQLARYL...GNTV.DLSCFDFFR.TGKMMPGKM							
BfMBD2	MERRRMDCPALPK	EWKREEVTRKSGLSAGKSDVYYYFSP.	SGKKFRSKPQLARYL...GDTF.DLSCFDFFR.TGKMLSSKV							
MmMB3	MERKRWECPALPQ	GWEREVPRRSGLSAGHRDVFYFSP.	SGKKFRSKPQLARYL...GCSM.DLSTFDFFR.TGKMLMNKM							
HsMBD3	MERKRWECPALPQ	GWEREVPRRSGLSAGHRDVFYFSP.	SGKKFRSKPQLARYL...GCSM.DLSTFDFFR.TGKMLMSKM							
DrMBDa	MERK.....	.....SP.	TGKKFRSKPQLVRYL...CKSM.DLSSFDFFR.TGKMLMSKL							
DrMBDb	~EKN.....	.....DP.	TGKKFRSKPQLARYL...CNSM.DLSSFDFFR.TGKMLMSKL							
AdMBD	VEKKKFECSALPK	GWOREEVIRKSGISAGKVDVYYYFSP.	NGKKFRSKPQLARYL...CDAL.DLATFDFFR.SGK.INSLL							
DvMBD	~~~~~	GWKREEVHKSGCNTG.VDVYFSP.	TGKRIESKPQLAROL...SAEMFVSNFCFQ.TGRMQIQ~							
DmMBD	TERKRVDCSVLPK	GWORDEV.RKSGSSAXKVDVYFSP.	TGKRAEGKPQ.....DIAIPDFQ.PGKMPHCAL							
LeMBD	~~~~~	HQDGWVIEEVPRTDGSS..MVDKYYYEPG	SGQKFRSRIAAQRYLAQMRENVPLSATLEELKENKPLSKMF							



C-terminal histidine tagged GST-PBM1-MBD-6H and GST-mPBM1-MBD-6H domains were eluted from the resin with 0.4M imidazole and aliquots from fractions were analysed by SDS-PAGE. The recombinant proteins were well expressed and readily detectable by Coomassie Blue dye staining. The size of the protein bands are in agreement with the predicted molecular weights of GST-PBM1-MBD-6H (42.3kDa) (see Figure 3.14A) and GST-mPBM1-MBD-6H domains (46.9kDa).

The 5' end of the MBD of mPBM1 is alternatively spliced in the mPBM1c isoform (see Figure 3.7). To test whether the presence or absence of this alternatively spliced region influences the MBD of mPBM1 in the recognition of DNA this minimal MBD was expressed in bacteria as a GST-fusion either containing or not the 95 amino acids encoded by the 285bp mPBM1c splice isoform. A 286bp cDNA encoding the minimal MBD of mPBM1 and a 616bp cDNA fragment encoding the MBD plus the 285bp splice variant present in the mPBM1c isoform were PCR amplified with oligonucleotides containing 5' BamHI and 3' EcoRI sites. These fragments were restriction digested and ligated to BamHI/EcoRI linearised pGEH expression vector to generate the expression constructs pGEH(-285)mPBM1-MBD and pGEH(+285)mPBM1-MBD :

m2K1-285/F	: 5'-GGTACAGGATCCCGAAGAAGAGTGGCAGAC-3'
m2K1-285/R	: 5'-TGGATTGAATTCTTGCTCTGGGTCTATCTG-3'
+285BP-m2K1MB	: 5'-GATGAAGGATCCGATGAAGATGATGAGTC-3'



**Figure 3.14 : Purification of recombinant proteins by affinity chromatography**

**A. Purification of the GST-PBM-1-MBD-6H**

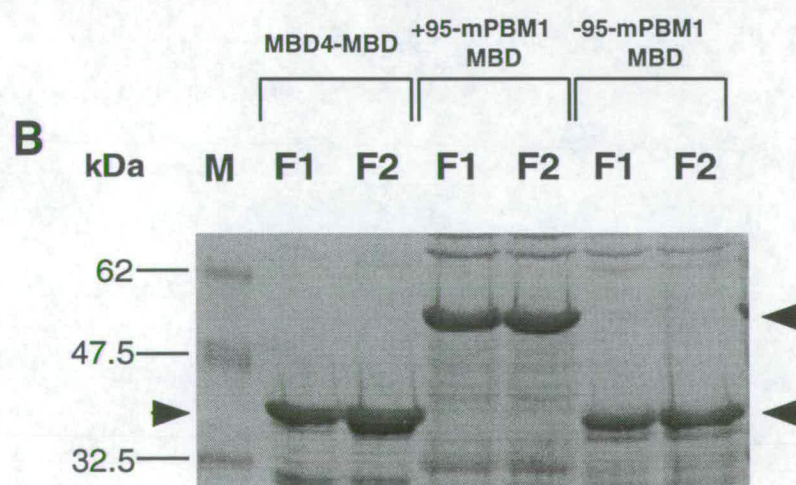
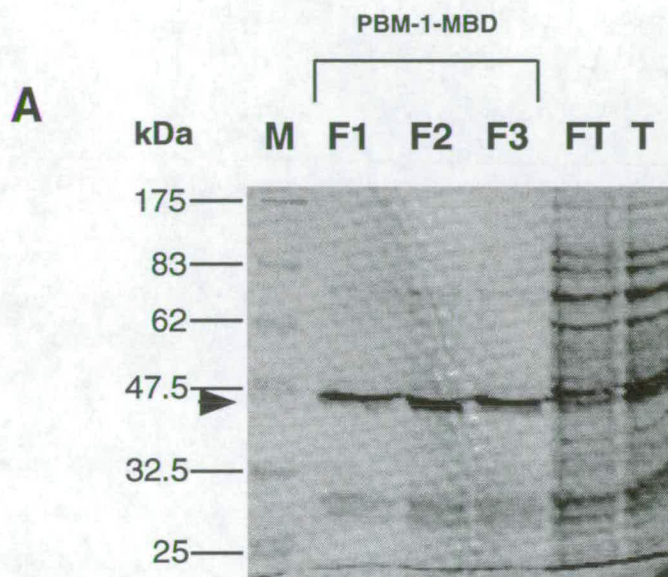
BL21(DE3)pLysS bacteria containing the pGEH-PBM-1-MBD construct were grown to OD<sub>600</sub> of 0.4 and induced with 0.4mM IPTG for 3 hours. Extracts from these cells were passed over a nickel agarose column to adsorb the C-terminal histidine tagged-PBM-1 protein. T is total proteins loaded onto the column and FT is the flow through from the column during washing steps to remove non-specifically bound proteins. F1, F2 and F3 are fractions one to three eluting from the column following addition of 0.4M imidazole. Aliquots were separated on 10% polyacrylamide gels. A band of the predicted size of GST-PBM-1-MBD-6H (42.3kDa) is indicated.

**B. Purification of 1. the GST-mMBD4-MBD-6H 2. GST-(+95)mPBM1-MBD-6H and 3. GST-(-95)mPBM1-MBD-6H.**

1. Bacterial cells containing the pGEH-PBM4-MBD plasmid were grown until the OD<sub>600</sub> was 0.4 and protein expression was induced with 0.4mM IPTG at 37°C for 3 hrs. Bacterial extracts were incubated with GSH-Sepharose beads to bind the GST-tagged mMBD4-MBD, washed extensively to remove non-specifically bound proteins and finally eluted by addition of 15 mM reduced glutathione (rGSH). F1 and F2 are fractions 1 and 2 collected following final elution. Aliquots from these fractions were analysed by SDS-PAGE. A band in agreement with the expected size of 37kDa for the GST-mMBD4-MBD-6H is indicated.

2. Bacteria were transformed with pGEH(+285)mPBM1-MBD to overexpress the mPBM1-MBD containing the 95 amino acid N-terminal splice variant encoded by the mPBM1c isoform (see section 3.2.4). Extracts from these cells were treated exactly as in Figure 3.15A. Fractions 1 and 2 contained a band of the predicted size for the GST-(+95)mPBM1-MBD-6H fusion protein, 49.1kDa.

3. Bacteria transformed with pGEH(-285)mPBM1-MBD were treated as in Figure 3.14 A above for purification of GST-(-95)mPBM1-MBD-6H. Fractions 1 and 2 contained a band of the predicted size of 37.9kDa.





These constructs were electrotransformed into *E. coli* BL21 (DE3) pLysS, grown in LB media until OD<sub>600</sub> was 0.4 and protein expression was induced with 0.4mM IPTG at 37°C for 3 hrs. The recombinant proteins were purified by nickel agarose (Qiagen) as described previously and were analysed by SDS-PAGE. The recombinant proteins were expressed well as judged by Coomassie blue staining and the size of the protein bands agree with the predicted molecular weights of the GST-(-95)mPBM1-MBD-6H (37.9 kDa) and the GST-(+95)mPBM1-MBD-6H domains (49.1 kDa) (see Figure 3.14B).

The MBD of MBD4 shows a reduced homology to the MBD of MeCP2 similar to the MBD of the *C. elegans* PBM-1 and mouse PBM1 proteins (see Figure 3.13). Hendrich et al., (1999) have demonstrated that the MBD of MBD4 is capable of binding to with greater affinity to m<sup>5</sup>CpG/TpG mismatches than to probes containing either symmetrically methylated DNA or CpG/TpG mismatches. In order to use the minimal MBD of the mouse MBD4 protein as a positive control in DNA binding studies, a 278bp fragment encoding the MBD was PCR amplified with primers containing 5' BamHI and 3' EcoRI sites. This fragment was subcloned to BamHI/EcoRI linearised pGEH expression vector to generate the expression construct pGEH-PBM4-MBD :

mMBD4/F : 5'-CTGCTGGATCCCTGTCTAGTACCACAGCG-3'

mMBD4/R : 5'-AGCTGGAATTCTTTGGTGTTTATAACCGGG -3'



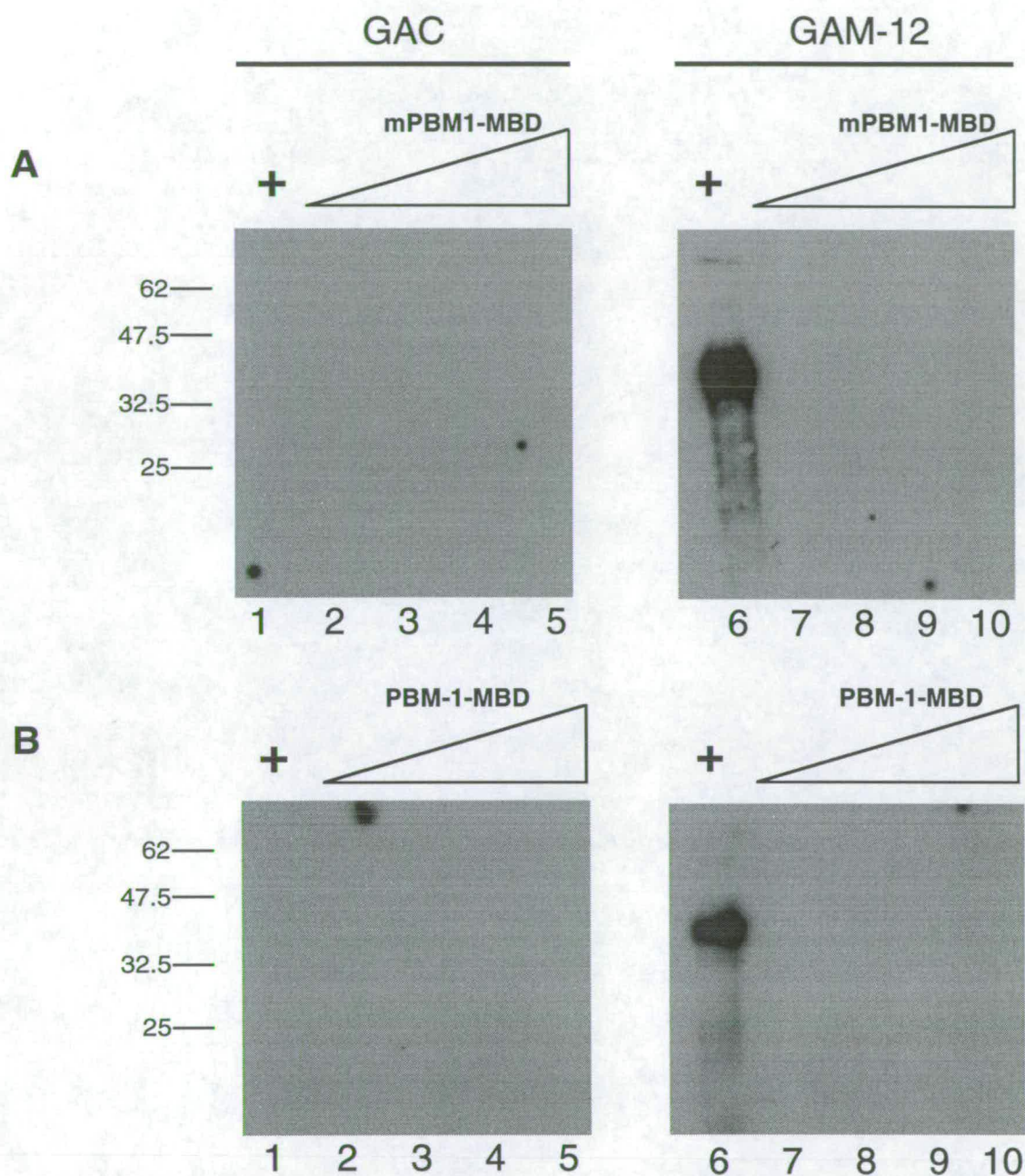
The pGEH-PBM4-MBD construct was electrotransformed into *E. coli* BL21 (DE3) pLysS cells, grown in LB media until OD<sub>600</sub> was 0.4 and protein expression was induced with 0.4mM IPTG at 37°C for 3 hrs. Following induction, the pGEH-PBM4-MBD construct generates a fusion protein with an N-terminal tagged glutathione S-transferase (GST) fused in-frame to the MBD of MBD4 and a C-terminal tag of six histidine residues. The MBD of MBD4 was purified on glutathione-Sephadex as follows. The induced bacteria were pelleted by centrifugation at 4K for 15 min at 0°C and then washed twice in GST-lysis buffer (50mM Hepes Tris.HCl pH 7.9, 0.15M NaCl, 10% glycerol, 1mM EDTA). The pellets were lysed in a buffer containing protease inhibitors to minimise proteolytic degradation and high salt (0.15M NaCl) to facilitate the solubilisation of the recombinant protein. The bacterial cells were lysed and the extract was incubated with GSH-Sepharose beads to bind the GST-tagged mMBD4-MBD, washed extensively to remove non-specifically bound proteins and finally eluted by addition of 15 mM reduced glutathione (rGSH). Aliquots from these fractions were analysed by SDS-PAGE and the recombinant MBD of mouse MBD4 was expressed well as shown by Coomassie Blue staining. The size of the protein band is in agreement with the predicted molecular weight of GST-mMBD4-MBD-6H (37.0kDa) (see Figure 3.14B).

### 3.2.7 Analysis of the methyl-CpG binding and DNA binding activity of the *C. elegans* PBM-1 MBD and mouse PBM1 MBDs

In order to characterise sequence specific DNA-binding proteins with affinity for the promoter of the transferrin receptor gene, Miskimins et al., (1985) developed the southwestern assay. In essence, a southwestern assay involves proteins (either cellular or recombinant) separated by SDS-PAGE being immobilised onto nitrocellulose membranes by electroblotting and then incubated with  $^{32}\text{P}$ -labelled DNA probes containing sequences of interest.

To determine whether the GST-PBM-1-MBD-6H and GST-mPBM1-MBD-6H domains were capable of binding to symmetrically methylated DNA duplexes a southwestern analysis was performed. Increasing amounts of these recombinant proteins and a positive control, the MBD of MeCP2, were separated by SDS-PAGE, electroblotted onto nitrocellulose membranes and challenged with either unmethylated (GAC) or methylated (GAM12)  $^{32}\text{P}$ -labelled DNA probes in the presence of *E. coli* competitor DNA. These southwestern assays showed that the 85 amino acid MBD of MeCP2 specifically bound to a multiply methylated DNA duplex but had no detectable affinity for unmethylated DNA probes as previously demonstrated by Nan et al., (1993). However, by Southwestern analysis the MBD of the *C. elegans* PBM-1 and mouse PBM1 proteins showed no demonstrable affinity for either the methylated (GAM12) or unmethylated (GAC) probes in the presence of *E. coli* competitor DNA (Figure 3.15).





**Figure 3.15 :** Determination of the affinity of mPBM1-MBD and PBM-1-MBD for methylated DNA by Southwestern analysis

A. Lane 1 : 50ng of recombinant MeCP2 MBD, Lanes 2-5 : 50, 100 and 150ng of recombinant mPBM1-MBD immobilised on a nitrocellulose filter were assayed for binding to unmethylated (GAC)-<sup>32</sup>P labeled probe in the presence of *E. coli* competitor DNA. Lane 6 : 50ng of recombinant MeCP2 MBD, Lanes 7-10 : 50, 100 and 150ng of recombinant mPBM1-MBD immobilised on a nitrocellulose filter was assayed for binding to methylated (GAM-12)-<sup>32</sup>P labeled probe in the presence of *E. coli* competitor DNA.

B. Recombinant PBM-1-MBD was assayed for binding to methylated (GAM-12)-<sup>32</sup>P labeled probe exactly as described for mPBM1.

An alternative assay to characterise the DNA-binding properties of proteins is the bandshift assay. A bandshift assay involves the incubation of recombinant proteins or nuclear extracts with radiolabelled nucleic acid probes containing sequences of interest and the subsequent electrophoretic separation of protein-DNA complexes (Strauss and Varshavsky, 1984).

In order to test whether the MBD from *C. elegans* PBM-1 and mouse PBM1 proteins have any discernible affinity for methylated template by an alternative method, a series of bandshift assays were performed. Recombinant GST-PBM-1-MBD-6H and GST-mPBM1-MBD-6H were incubated with end-labeled GAC (unmethylated) or GAM12 (methylated) probes (see Figure 3.18), in the presence of *E. coli* competitor DNA and the mixes run on 1.5% agarose gels to resolve DNA-protein complexes from free probe. Figure 3.16 shows that the MBD region of MBD1 preferentially binds to methylated DNA whereas no detectable affinity for unmethylated DNA was detected in this assay. In support of the southwestern analysis, intact recombinant proteins encompassing the MBD of *C. elegans* PBM-1 and mouse PBM1 did not bind to either methylated or unmethylated probes by bandshift analysis. Thus, by two independent techniques the MBD of the *C. elegans* PBM-1 and mouse PBM1 proteins do not form specific complexes with methylated probes, strongly suggesting that symmetric m<sup>5</sup>CpG sites are not the target for these domains.



**Figure 3.16 :** Determination of the affinity of mPBM1-MBD and PBM-1-MBD for methylated DNA by bandshift analysis

A. Lane 1 : 50ng of the MBD of human MBD1 incubated with unmethylated (GAC)-<sup>32</sup>P labelled probe in the presence of 2μg *E. coli* competitor DNA.

Lanes 2-4 : 100ng of recombinant mPBM1-MBD incubated with unmethylated (GAC)-<sup>32</sup>P labelled probe in the presence of 0, 1 and 2μg *E. coli* competitor DNA .

Lanes 5-7 : 100ng of recombinant mPBM1-MBD incubated with methylated (GAM-12)-<sup>32</sup>P labelled probe in the presence of 0, 1 and 2μg *E. coli* competitor DNA.

Lanes 8-10 : 50, 100 and 150ng of recombinant mPBM1-MBD incubated with unmethylated (GAC)-<sup>32</sup>P labelled probe in the presence of no competitor DNA.

Lane 11 : 50ng of the MBD of human MBD1 incubated with methylated (GAM12)-<sup>32</sup>P labelled probe in the presence of 2μg *E. coli* competitor DNA.

Lanes 12-14 : 50, 100 and 150ng of recombinant mPBM1-MBD incubated with methylated (GAM-12)-<sup>32</sup>P labelled probe in the presence of no *E. coli* competitor DNA.

Mixes were electrophoresed through a 1.5% agarose gel, dried onto DE81 filter paper and exposed to X-ray film.

B. Recombinant PBM-1-MBD was assayed for binding to methylated (GAM-12) labeled probe exactly as described for mPBM1.

GAC

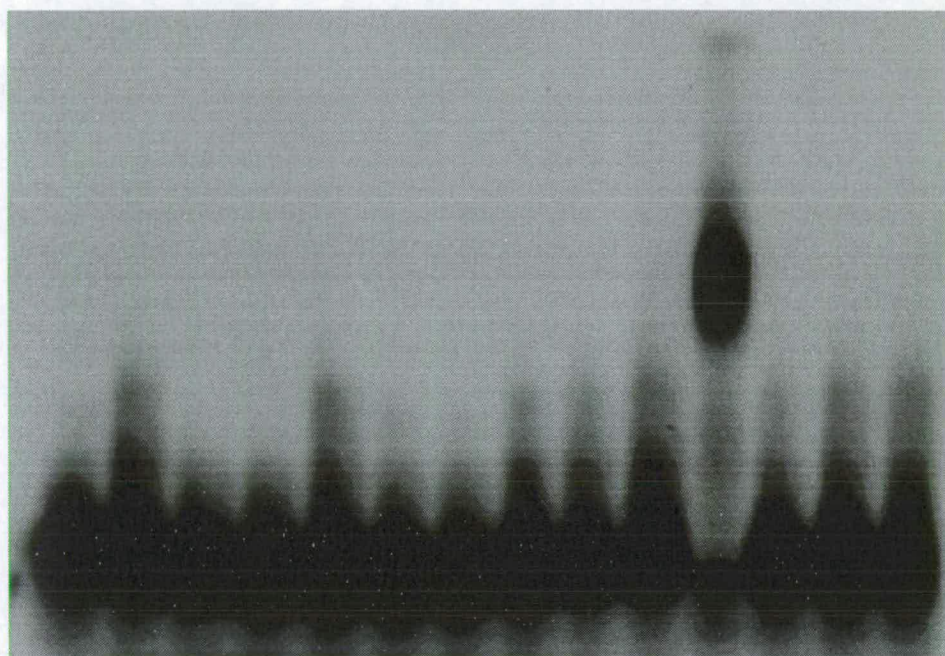
GAM-12

GAC

GAM-12

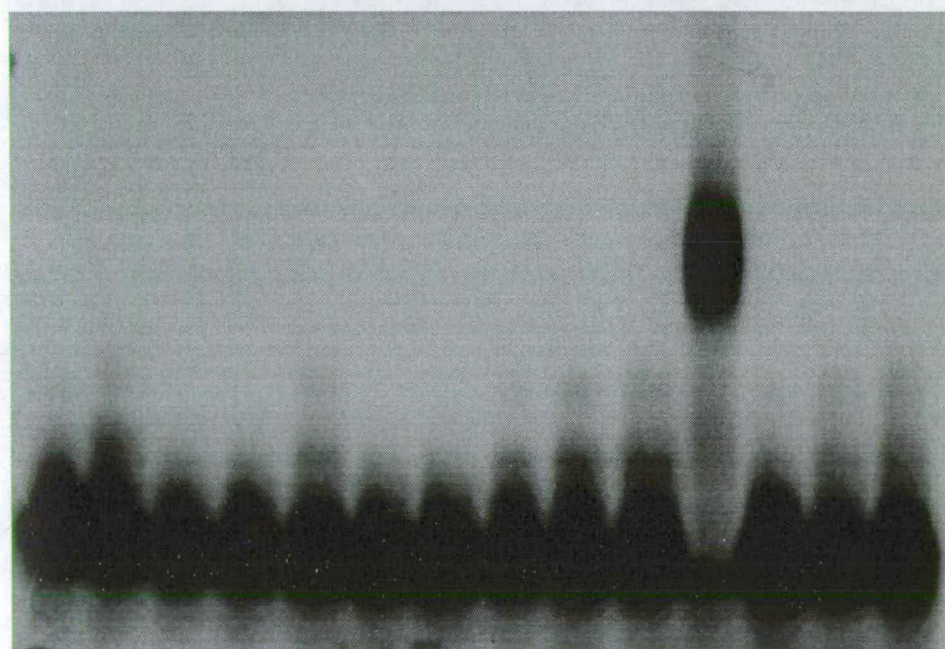
**A**

Free Probe



**B**

Free Probe



1 2 3 4 5 6 7 8 9 10 11 12 13 14



MBD4 binds with greater affinity to m<sup>5</sup>CpG/TpG mismatches than to probes containing symmetrically methylated DNA (m<sup>5</sup>CpG/m<sup>5</sup>CpG) or CpG/TpG mismatches (Hendrich et al, 1999). Thus both the DNA-binding MBD regions of MeCP2 and MBD4 are capable of binding to target sequences with additional but differing specificities. This raised the possibility that the *C. elegans* PBM-1 and mPBM1 proteins could bind to DNA target sequences other than symmetrically methylated DNA via their MBD.

The 5' end of the MBD of mPBM1 is alternatively spliced in the mPBM1c isoform. To test whether the presence or absence of this alternatively spliced region influences the recognition of DNA, the MBD of mPBM1 was expressed in bacteria as a recombinant fusion either containing, (+95)mPBM1-MBD or not, (-95)mPBM1-MBD the 95 amino acids encoded by the 285bp mPBM1c splice isoform. Figure 3.17 shows that the MBD of MBD4 binds preferentially to a DNA fragment containing m<sup>5</sup>C in a variety of sequence contexts (see probe in Figure 3.19). The MBD of MBD4 also weakly shifted an unmethylated DNA probe which is potentially attributable to the low affinity of this region of the MBD4 protein for target sequences other than symmetrically methylated DNAs or m<sup>5</sup>pG/TpG (Hendrich et al., 1999). The isoforms of the MBD of mPBM1 were tested for binding to these methylated and unmethylated probes. No specific complexes between either isoforms of the mPBM1-MBD and the methylated probe was observed as addition of the competitor poly (dI-dC) abolished any weak complexes

**Figure 3.17 :** Determination of the affinity of the (-95)-mPBM1-MBD and (+95)-mPBM1-MBD isoforms for methylated DNA by bandshift analysis

Lanes 1-2 : 100ng of recombinant MBD4-MBD incubated with the AtMBD-<sup>32</sup>P labelled probe unmethylated at C in the absence (-) or presence (+) of 1μg of poly(dI-dC) competitor DNA.

Lanes 3-4 : 100ng of recombinant (-95)-mPBM1-MBD incubated with the AtMBD-<sup>32</sup>P labelled probe unmethylated at C (see Figure 3.19) in the absence (-) or presence (+) of 1μg of poly(dI-dC) competitor DNA.

Lanes 5-6 : 100ng of recombinant (+95)-mPBM1-MBD incubated with the AtMBD-<sup>32</sup>P labelled probe unmethylated at C in the absence (-) or presence (+) of 1μg of poly(dI-dC) competitor DNA.

Lanes 7-8 : 100ng of recombinant MBD4-MBD incubated with the AtMBD-<sup>32</sup>P labelled probe methylated at C in the absence (-) or presence (+) of 1μg of poly(dI-dC) competitor DNA.

Lanes 9-10 : 100ng of recombinant (-95)-mPBM1-MBD incubated with the AtMBD-<sup>32</sup>P labelled probe methylated at C in the absence (-) or presence (+) of 1μg of poly(dI-dC) competitor DNA.

Lanes 11-12 : 100ng of recombinant (+95)-mPBM1-MBD incubated with the AtMBD-<sup>32</sup>P labelled probe methylated at C in the absence (-) or presence (+) of 1μg of poly(dI-dC) competitor DNA.



Unmethylated

Methylated

Poly (dl-dC)

-

+

-

+

-

+

-

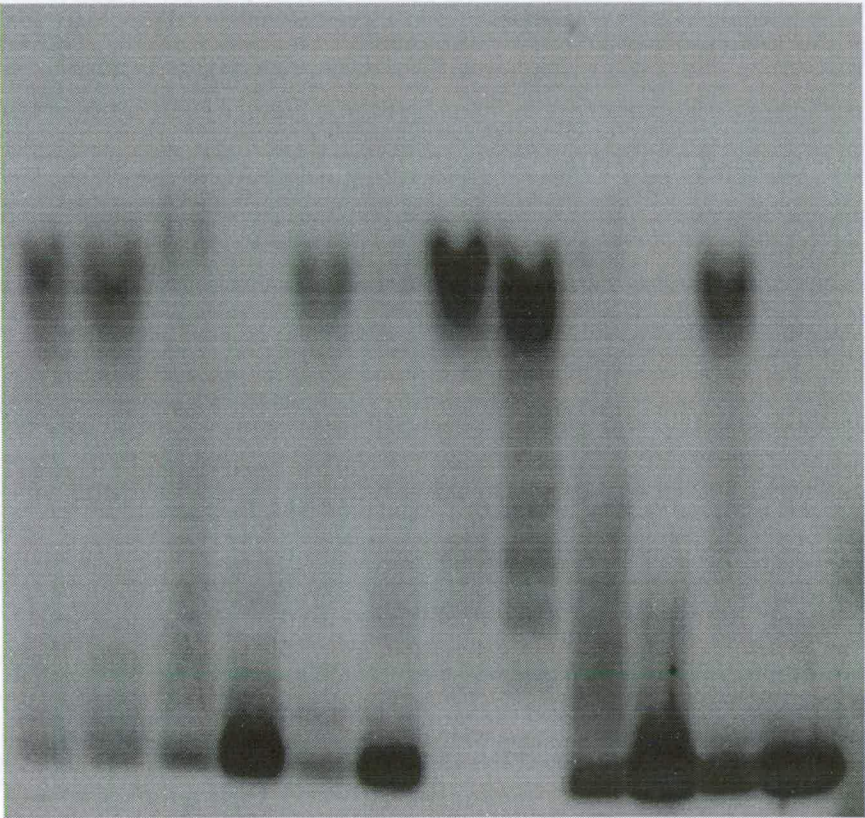
+

-

+

-

+



1

2

3

4

5

6

7

8

9

10

11

12

MBD4

-95

+95

MBD4

-95

+95

GAM-12	GAT <b>M</b> GAMGAMGAMGAMGAMGAMGAMGAMGAMGAMGAMG G <b>M</b> TG <b>M</b> TG <b>M</b> TG <b>M</b> TG <b>M</b> TG <b>M</b> TG <b>M</b> TG <b>M</b> TG <b>M</b> TG <b>M</b> TG <b>M</b> TAG	CA/GM-1	TCAGATTCGCGC <b>CA</b> GCTGCGATAAGCT AGTCTAAGCGCG <b>G</b> CGATGCTATTCTGA
GAC	GATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA CGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTTAG	CA/GT1	TCAGATTCGCGC <b>CA</b> GCTGCGATAAGCT AGTCTAAGCGCG <b>G</b> TGATGCTATTCTGA
MG/GT1	TCAGATTCGCGC <b>MG</b> GCTGCGATAAGCT AGTCTAAGCGCG <b>GT</b> CGATGCTATTCTGA	CA/GC1	TCAGATTCGCGC <b>CA</b> GCTGCGATAAGCT AGTCTAAGCGCG <b>G</b> CCGATGCTATTCTGA
MG/GC-1	TCAGATTCGCGC <b>MG</b> GCTGCGATAAGCT AGTCTAAGCGCG <b>G</b> CCGATGCTATTCTGA	CG/GT1	TCAGATTCGCGC <b>CG</b> GCTGCGATAAGCT AGTCTAAGCGCG <b>GT</b> CGATGCTATTCTGA

**Figure 3.18** : Probes used to test DNA-binding properties of the MBDs of PBM-1 and mPBM1

The letter **M** denotes m<sup>5</sup>C incorporated at this position of the oligonucleotide. The GAM-12 probe is composed of a symmetrical repeat of 12 m<sup>5</sup>CGA. The GAC probe is identical to GAM12 except C is substituted for m<sup>5</sup>C. The MG/GC-1 probe is a hemi-methylated DNA and the remaining probes contain internal mismatches (in bold).



**m<sup>5</sup>C-AtMBD**

5' MGGGAT**MM**ATGGAAA**AM**AMAGAMGAG**MT**TGT**MTMM**ATTGA  
 GMT**AMM**AG**MTMM**AG**MTT**MATGGAAGAA**MT**GTTTTAT**MM**GAA  
 AAGAG**MM**GGT**AMTMM**GAGAAAGAMGGAGATTGTGTTTGTGG**M**  
 T**MM**A**AM**GGGTGAAGAGATTAG**MTM**G**M**GGAAG**M**AGTTGGAG**M**A  
 GT**AMMT**GAAGG**MMM**AT**MMT**GG**MAATMMT**GT**MTMT**GAGTT  
 TGAGTGG**AMA****MT**GGGGAA**MTMM**AAGGAGGGAATT**M**GG3'

**AtMBD**

5' CGGGATCCATGGAAAACACAGACGAGCTTGTCTCCATTGA  
 GCTACCAGCTCCAGCTTCATGGAAGAACTGTTTTATCCGAA  
 AAGAGCCGGTACTCCGAGAAAGACGGAGATTGTGTTTGTGGC  
 TCCAACGGGTGAAGAGATTAGCTCGCGGAAGCAGTTGGAGCA  
 GTACCTGAAGGCCCATCCTGGCAATCCTGTCATCTCTGAGTT  
 TGAGTGGACAAC**T**GGGGAA**ACT**CCAAGGAGGGAATT**CC**GG3'

**Figure 3.19 :** The single stranded AtMBD fragment was used to generate a double stranded probe to test the affinity of the (-95)-mPBM1-MBD and (+95)-mPBM1-MBD isoforms for methylated DNA.

The AtMBD probe is a 247bp DNA fragment containing 55 cytosines in a variety of sequence contexts. Incorporation of m<sup>5</sup>C was carried out by PCR amplification of this 0.24kb using m<sup>5</sup>C substituted for C. The letter **M** denotes m<sup>5</sup>C incorporated at this position of the fragment. This fragment was annealed together to generate a double stranded DNA fragment for in vitro binding studies.

with these probes which were thus adjudged to be the result of non-specific interactions. Furthermore, in bandshift assays with probes containing hemimethylated DNA, CpG/TpG and m<sup>5</sup>pG/TpG mismatch probes no affinity for these sequences was observed with the MBD of either the *C. elegans* PBM-1 or mouse PBM1 proteins (data not shown).

### 3.3 Conclusions

A *C. elegans* gene *pbm-1* has been identified on the basis of homology to the methyl-CpG binding domain of the MBD protein family. The *pbm-1* gene has 8 exons spanning 5191bp of genomic DNA on chromosome III. A 4173bp cDNA clone derived from this locus contained an open reading frame predicted to encode a large protein of 1390 amino acids. In addition, to the MBD homology BLASTp searches revealed the existence at the extreme C-terminus of a bipartite PHD-zinc (C4HC3) finger and an adjacent bromodomain. Intriguingly, these motifs have been found in a diverse array of chromatin-associated transcriptional regulators.

A mammalian ortholog of the *C. elegans* PBM-1 protein was identified by screening a mouse brain cDNA library and potentially encoded a protein with an N-terminal MBD and a C-terminal PHD-zinc finger/bromodomain region. This murine cDNA was named mPBM1. The mPBM1 cDNA was the product of a complex series of alternative splicing events and RT-PCR analysis detected a 95 amino acid splicing



event immediately N-terminal of the MBD. Both spliced and unspliced forms of the message were detected in a range of somatic and tissue cultured cells.

During the isolation of the mPBM1 cDNA sequence a partial murine cDNA, distinct from mPBM1, was identified that was highly homologous to the *C. elegans* PBM-1 PHD-zinc finger/bromodomain region. This cDNA was named mPBM2. In addition, corresponding paralogs hPBM1 and hPBM2 were identified in humans. This suggests the possible existence of a family of PHD-zinc finger, bromodomain and MBD (PBM)-containing proteins in mammals.

Recombinant proteins comprising the MBD of the *C. elegans* PBM-1 and mouse PBM1 proteins were generated and their affinity for methylated DNA was tested. By two independent approaches, no detectable binding of the MBD of the *C. elegans* PBM-1 and mouse PBM1 proteins to a methylated template was observed whereas the MBDs of MeCP2 and MBD1 bound methylated probes specifically. In addition, the presence of a 95 amino acid region N-terminal to the MBD of mPBM1 derived from a predicted alternative splicing event was not found to target the MBD region to methylated DNA duplexes.

## **CHAPTER FOUR: LOCALISATION AND TRANSCRIPTIONAL REGULATORY PROPERTIES OF THE mPBM1 PROTEIN**



## **4.1 Introduction to mPBM1 localisation**

Green Fluorescent Protein (GFP) has attracted enormous interest as a biological marker for protein localisation and gene expression studies since the initial reports that the transcription of the cloned gene generated strong visible fluorescence in a range of cell types (Chalfie et al., 1994; Inouye and Tsuji, 1994). GFP has greatly aided the analysis of localisation patterns in the absence of immunofluorescence, however it does suffer from the caveat that expression levels are often not reflective of endogenous promoter strength and interpretation of localisation patterns must reflect this (Cubitt et al., 1995). However, translational fusions between mammalian proteins and reporter molecules have been found to localise to sites where the endogenous protein is known to reside, strongly suggesting that the localisation pattern discerned with these translational fusions is biologically meaningful (Gossler et al., 1989).

## **4.2 Results**

### **4.2.1 Nuclear localisation of a full length GFP-mPBM1 fusion**

To investigate the subcellular distribution of mPBM1 and specifically determine whether the protein was nuclear or cytoplasmically localised, the entire mPBM1 ORF was fused in-frame to the GFP cDNA sub-cloned downstream of the strong cytomegalovirus (CMV) promoter and transfected into mouse EFS2 fibroblast cells for transient expression. This expression construct was generated in two steps. The first manipulation involved the PCR amplification of an 1876bp fragment, comprising the extreme 3' end of the mPBM1 ORF. This fragment contained an internal BamHI

site (from the mPBM1 cDNA) and incorporated a 3' NheI site for subcloning in-frame into a BamHI / NheI linearised pCMX-GFP2 vector (Hendrich and Bird, 1998). This clone is termed pCMX-GFP2-3'mPBM1. A 4986bp fragment comprising the 5' end of mPBM1 was amplified by PCR. This fragment contained the internal BamHI site from the mPBM1 cDNA and incorporated a 5' BamHI for subcloning into the BamHI linearised pCMX-GFP2-3'mPBM1. This construct is termed pCMX-GFP2-mPBM1. These clones were restriction mapped to identify correctly oriented inserts and all clones in this study were sequenced to ensure no PCR-generated mutations were present.

Bam2K1/F	: 5'-GGATGGTGGAGAATCATTGACC -3'
NheI GFP2/R	: 5'-GAGATTGCTAGCTCAGCTCACTTTGAAGG-3';
GFP Bam ATG/F	: 5'-TTATACGGATCCATGGAGTCTGGAGAACTG-3'
Bam2K1/R	: 5'-CTTGTTCTGACACAGGCTTGG-3'

A range of concentrations (0.5µg - 10µg) of this full length mPBM1-GFP fusion were transfected into mouse EFS2 fibroblasts by overlaying 40% confluent cells with a DNA-liposome mix as described in section 2.2.5. Ectopically expressed pCMX-GFP2-mPBM1 was nuclear localised in murine fibroblasts (see Figure 4.1). Expression in EFS2 cells of full length mPBM1-GFP fusion gave a prominent, uniformly distributed pattern of fluorescent foci throughout the nucleoplasm with the exception of the nucleoli from which it appeared to be absolutely excluded (Figure 4.1A). Numerous localisation studies on MeCPs have taken advantage of the heavily methylated mouse major satellite repeat which contains some 50% of all m<sup>5</sup>C in the



**Figure 4.1 : Sub-cellular localisation of the full length mPBM1 protein**

**A. mPBM1-GFP fusions are nuclear localised in murine fibroblasts**

Transient expression in EFS2 cells of a full length mPBM1-GFP fusion, pCMX-GFP2-mPBM1 showed a granular pattern of fluorescent foci within the nucleus and absolute exclusion from the nucleoli. The DAPI bright spots are known to contain a large fraction of m<sup>5</sup>C in the mouse genome (Miller et al., 1974). No co-localisation of mPBM1-GFP fusions with DAPI bright spots could be observed strongly suggesting that methylated regions are not the *in vivo* target.

**B. Full length mPBM1-GFP does not associate with chromosomes on entry to mitosis**

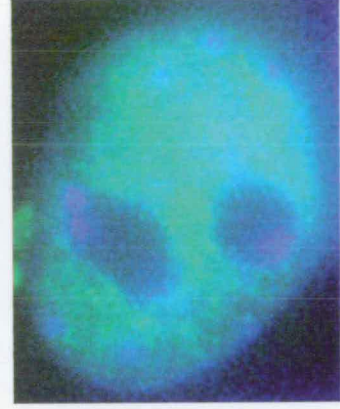
No-coincidence of the GFP and DAPI signals was observed in cells fixed at the prometaphase stage of the cell cycle.

**GFP**

**DAPI**

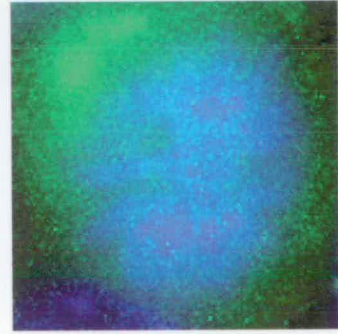
**MERGED**

**A**

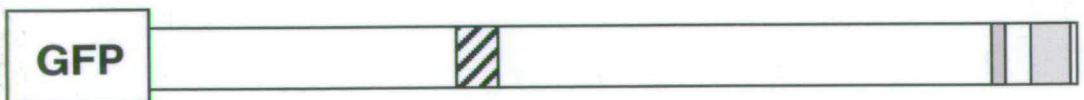


**INTERPHASE**

**B**



**PROMETAPHASE**



**MBD**

**PHD-BrD**



mouse genome (Miller et al., 1974). These regions of mouse major satellite that are constitutively heterochromatic within the mouse nucleus bind strongly to the DNA dyes Hoechst 33258 and DAPI (Miller et al., 1974). This has allowed investigators to use the major satellite repeat as an in vivo assay for association between an MeCP and heavily methylated heterochromatin. Nan et al., (1996) for example, found that transiently expressed MeCP2- $\beta$ geo fusions localised to pericentromeric heterochromatin and Hendrich and Bird (1998) found that MBD1 also co-localised with brightly staining DAPI foci of the mouse major satellite. In support of the in vitro DNA-binding studies that did not detect binding of the MBD of mPBM1 to a methylated template, no strong co-localisation of mPBM1-GFP fusions with DAPI bright spots could be observed strongly suggesting that methylated regions are not the in vivo target. Furthermore, full length mPBM1-GFP was not found to associate with chromosomes at the time of entry to mitosis, with no coincidence of the DAPI and GFP signals detectable in those cells fixed at this stage (Figure 4.1B).

#### **4.2.2 Punctate localisation of a GFP-mPBM1 $\Delta$ MBD**

Is the diffuse, punctate nuclear localisation pattern of full length mPBM1-GFP dependent on the N-terminal region containing the MBD ? To answer this question, a deletion construct was engineered that removed this region and was transfected ectopically into mouse ESF2s. This expression construct was generated by the PCR amplification of a 2826bp fragment comprising the mPBM1 cDNA downstream of the MBD of mPBM1, that incorporated a BamHI site and contained an internal

BamHI site (from the mPBM1 cDNA). This fragment was BamHI digested and subcloned in-frame into the BamHI linearised pCMX-GFP2-3'mPBM1 construct (the pCMX-GFP2 expression vector containing the 3' end of mPBM1). This clone is termed GFP-mPBM1 $\Delta$ MBD.

-MBD GFP/F : 5'-GTCATTGGATCCATCAGGGCAATGGACGGC-3'  
BamPBM1/R : 5'-CTTGTTCTGACACAGGCTTGG-3'

The diffuse localisation pattern of fluorescent foci seen with the full length mPBM1 fusion was recapitulated following overexpression in mouse EFS2 fibroblasts of the GFP-mPBM1 $\Delta$ MBD construct (Figures 4.2A and 4.2B). Transfection experiments carried out with 1.0 $\mu$ g, 2.5 $\mu$ g, 5.0 $\mu$ g and 10 $\mu$ g of GFP-mPBM1 $\Delta$ MBD led to similar observations. Thus, no domains reside within the N-terminal region of mPBM1 responsible for conferring either the nuclear importing or for mediating the distinct, granular pattern of localisation of the mPBM1 protein in mouse nuclei.



**Figure 4.2 :** Deletion of the MBD of mPBM1 does not alter the nuclear localisation pattern

A. The full length mPBM1-GFP localisation pattern in murine fibroblasts

B. Localisation of the mPBM1-GFP without the N-terminal MBD is not perturbed following expression of the deletion

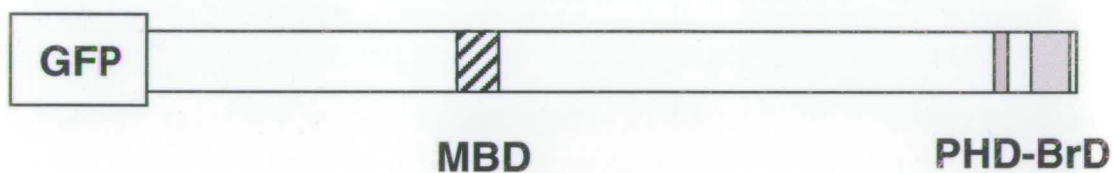
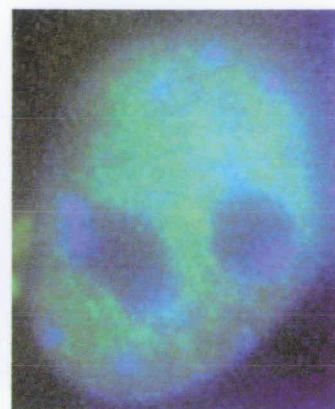
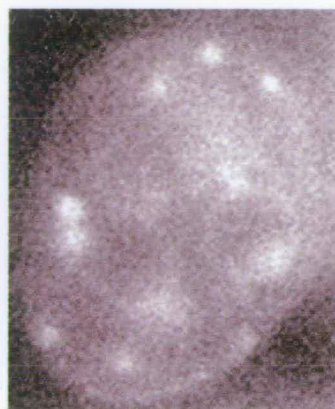
Following transfection of the GFP-mPBM1 $\Delta$ MBD expression construct the pattern of nuclear diffuse foci of signal observed with the full length mPBM1 fusion was observed. The deleted N-terminal region of mPBM1 therefore does not contain the nuclear localisation domain of mPBM1 or elements required for conferring the wild type diffuse localisation seen with the full length mPBM1 protein.

**GFP**

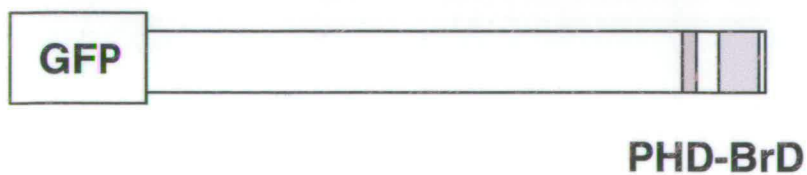
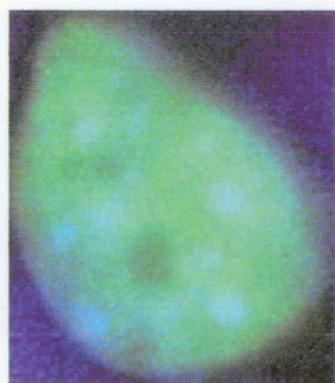
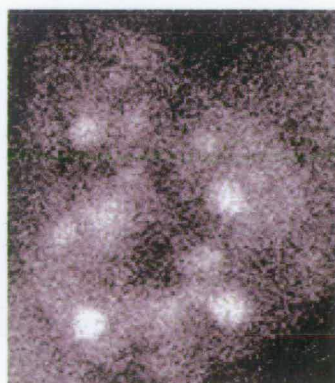
**DAPI**

**MERGED**

**A**



**B**





#### 4.2.3 Perturbed localisation of a GFP-mPBM1 $\Delta$ PHD/BrD

To determine whether the bipartite C-terminal PHD-zinc finger/bromodomain was responsible for this distinct nuclear localisation pattern, an expression construct was engineered that removed this region. This expression construct was generated by the PCR amplification of a 4986bp fragment comprising the 5' end of the mPBM1 cDNA including the putative initiator methionine and contained an internal BamHI site from the mPBM1 cDNA upstream of the bipartite PHD-zinc finger and bromodomain. This fragment was BamHI digested to delete the PHD-zinc finger bromodomain motifs and subcloned in-frame into the BamHI linearised pCMX-GFP2. This construct was termed mPBM1 $\Delta$ PHD/BrD :

GFP ATG/F : 5'-TTATACGGATCCATGGAGTCTGGAGAACTG-3'  
BamPBM1/R : 5'-CTTGTTCTGACACAGGCTTGG-3'

Strikingly, the diffuse, granular GFP signal seen within the nucleus with full length mPBM1 was abolished following deletion of the C-terminal region of the protein (compare Figure 4.3A and Figure 4.3B). The nuclear localisation pattern with this deletion construct consisted of intensely labelled foci that localised in discrete nuclear domains in marked contrast to the wild type protein and the fusion lacking the MBD. This clear difference in localisation patterns between the wild type and C-terminally truncated proteins was importantly observed in independent transfection experiments. The truncated proteins still retain the ability to localise to the nucleus, compatible with the fusions still retaining a functional nuclear localisation signal.

**Figure 4.3 :** Deletion of the PHD-zinc finger and bromodomain of mPBM1 results in an altered sub-nuclear localisation pattern

A. The full length mPBM1-GFP localisation pattern in murine fibroblasts

B. Perturbed localisation of the mPBM1-GFP fusion lacking the C-terminal PHD-zinc finger and bromodomain regions

Transient expression of the GFP-mPBM1 $\Delta$ PHD/BrD construct in mouse EFS2 cells abolished the wild type localisation pattern and the truncated fusion localised in large aggregates of intensely labelled foci. The mPBM1 $\Delta$ PHD/BrD is still nuclear localised, consistent with the fusions still retaining a functional nuclear localisation signal. Thus, mPBM1 requires a C-terminal element for spatial organisation of the protein in the nucleus.

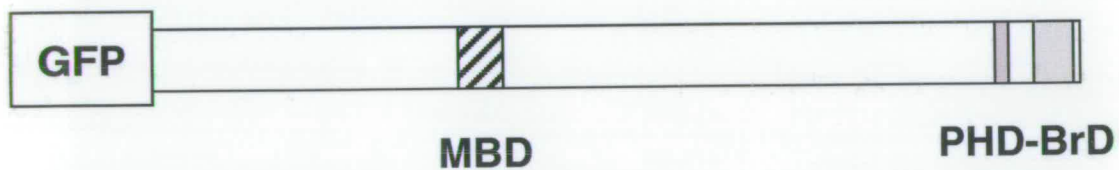
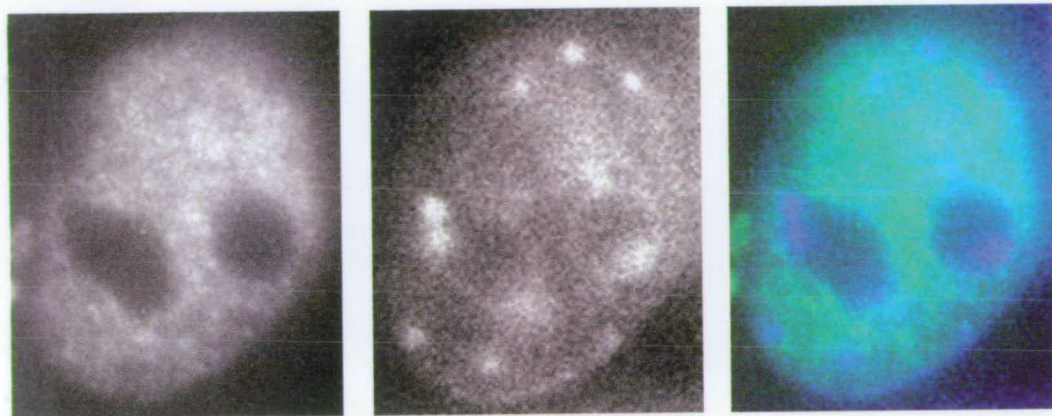


**GFP**

**DAPI**

**MERGED**

**A**



**B**



Thus, the deletion mutant lacking the C-terminus containing the PHD zinc finger and bromodomain lacks an element that is responsible for conferring the wild type localisation pattern. Site-directed mutagenesis or deletion mapping would determine the precise region responsible for the wild type nuclear localisation of the mPBM1 protein.

### **4.3 Conclusions**

This section describes the localisation of an mPBM1-GFP fusion protein in mouse EFS2 cells. Using translational fusions with a GFP reporter molecule, the full length mPBM1 protein was found to be uniformly distributed in the nucleus of mouse fibroblasts. No co-localisation of the full length mPBM1 GFP signal with the heavily methylated mouse major satellite DAPI signal was observed. This strongly suggests that full length mPBM1 does not associate with methylated domains in vivo, consistent with in vitro binding studies with the MBD of mPBM1, however these findings do not formally exclude the possibility that mPBM1 associates with a DNA target other than methyl-CpG rich domains.

The full length mPBM1 protein was not found to associate with chromosomes at the time of entry to mitosis as no co-localisation of the GFP and DAPI signal was observed at this time. Deletion of the N-terminus of mPBM1 containing the MBD did not alter the sub-nuclear localisation pattern demonstrating that no elements reside within this region responsible for the wild type localisation of the protein. The

localisation in a uniformly distributed pattern in the nucleoplasm does however require the extreme C-terminus of mPBM1 which contains the PHD-zinc finger and bromodomain motifs. Deletion of these motifs results in a strikingly different distribution pattern with the GFP signal present in discrete foci of intensely labelled fluorescence. A C-terminal element in mPBM1 is thus responsible for the spatial organisation of the mPBM1 protein within the nucleus of mouse fibroblasts.

#### 4.4 Transcriptional regulatory properties of the mPBM1 protein

The mPBM1 protein contains a bipartite PHD-zinc finger and bromodomain motif at its extreme C-terminus previously found in a large number of nuclear regulatory or chromatin-associated proteins (Aasland et al., 1995; Jeanmougin et al., 1997). The molecular mechanisms underlying the regulation of the clustered homeotic genes (HOM-C) in *Drosophila* is well studied (Paro, 1993). The finding that many of the proteins implicated in the regulation of HOM-C expression patterns possess PHD-zinc fingers and/or bromodomains serves to illustrate the central role of these domains in gene regulation at the level of the chromatin template. Transiently expressed regulatory proteins of the trithorax (trx-G) and polycomb (Pc-G) classes, that operate to activate or silence homeotic gene expression respectively, mediate their effects by altering chromatin structure of target genes. The repressive Pc-G activities are thought to silence HOM-C activity by the propagation of a heterochromatin-like structure across these clustered homeotic genes (Paro, 1990). Brahma is required for the activation of numerous homeotic loci during *Drosophila* development and loss-of-function Brahma mutations suppress mutations in *Polycomb* which functions to repress homeotic gene expression (Tamkun et al., 1992). Brahma possesses extensive homology, including a bromodomain to SWI2/SNF2 that operates as a transcriptional activator in *Saccharomyces cerevisiae* by modulating the chromatin structure of genes it regulates (Kruger et al., 1995). In addition to the bromodomain, the PHD-zinc finger motif has been found in several chromatin-associated proteins including the Trithorax and Polycomb-like proteins which exert antagonistic effects on expression

and chromatin structure during *Drosophila* development. That the PHD-zinc finger and bromodomain motifs are present in both transcriptional activators and repressors of the clustered homeotic genes suggests that both positive and negative chromatin-mediated transcriptional regulation might utilise similar mechanisms.

Further links between the possession of these motifs and gene regulation involving chromatin is suggested by their frequent occurrence of PHD-zinc finger and/or bromodomains in many of the proteins that modulate the acetylation status of the nucleosomal core histones. The intrinsic histone acetyltransferase, p300/CBP like the mPBM1 protein contains a C-terminal bromodomain and an adjacent PHD-zinc finger region. The finding that numerous other acetyltransferases also contain these motifs suggests that these domains are essential for the operation of these proteins within the environment of chromatin.

#### **4.5 Results**

To investigate whether mPBM1 was capable of modulating a reporter gene in vivo, the MBD at the N-terminus of mPBM1 was replaced with the DNA-targeting domain of the yeast transcription factor GAL4 in the pCMV-GAL4 expression vector. The ability of this GAL4-mPBM1 fusion driven by a cytomegalovirus promoter to affect a reporter gene containing GAL4 binding sites was assayed in mouse L929 cells. This GAL4-mPBM1 construct was created by PCR amplification of a 4161bp PCR product that incorporated 5' and 3' BclI sites for subcloning

directly into a BamHI linearised pCMV-GAL4 expression construct, generating the construct, pCMV-GAL4-mPBM1.

-MBDGAL4 mPBM1/F : 5'-TCATTCTGATCATCAGGGCAATGGACGGCC-3'  
pCMV-GAL4-mPBM1/R : 5'-GAGATTGATCATTGAGCTCACTTTGAAGGT-3'

mPBM1 transcripts are detectable in mouse L fibroblasts and thus putative mediators of any effect on transcription would reasonably be expected to be expressed within this cell line also (see Figure 3.6). The effect of a transiently expressed GAL4-mPBM1 fusion on the transcription of a  $\beta$ -galactosidase ( $\beta$ -gal) reporter gene driven by the human  $\beta$ -actin promoter was quantified in vivo by co-transfection of a range of increasing amounts of effector (25ng-500ng GAL4-mPBM1) with 2 $\mu$ g of reporter constructs that contained five GAL4 DNA-binding elements (p $\beta$ G5 BglII : G5) or did not (p $\beta$ geoN/B : G0). Using reporter constructs that were identical, except for the possession of upstream GAL4 binding elements allowed the effect of the GAL4-mPBM1 effector on  $\beta$ -gal activity to be determined following the specific tethering of the fusion protein at a promoter as opposed to any non-specific activity on the reporter gene.

Mouse L929 fibroblasts were co-transfected when 30% confluent with the pCMV-GAL4-mPBM1 fusion and either the p $\beta$ G5 BglII or p $\beta$ geoN/B reporter constructs by the DEAE-dextran method (section 2.2.5). Figure 4.4 shows that the GAL4-mPBM1 effector was capable of a modest dose-dependent and specific repression of the p $\beta$ G5 BglII construct whereas no repressive effect on the p $\beta$ geoN/B reporter

was observed. Transcription of the p $\beta$ G5 BglIII construct was reduced by >50% when co-transfected with 500ng of the pCMV-GAL4-mPBM1 fusion. In contrast, overexpression of the GAL4-mPBM1 fusion protein in L929 cells co-transfected with p $\beta$  $\beta$ geoN/B, that does not contain upstream GAL4 binding sites, lead to a robust transcriptional activation of reporter gene transcription. One potential explanation for this observation is that following the overexpression of the GAL4-mPBM1 fusion, there would result a titration of putative ancillary proteins required for silencing that would manifest itself as increased reporter gene activity in this assay. However, this non-specific effect of GAL4-mPBM1 on reporter gene transcription of p $\beta$  $\beta$ geoN/B does not result in a repression of transcription and thus the effect of GAL4-mPBM1 on p $\beta$ G5 BglIII is likely the consequence of the specific tethering at the upstream GAL4 binding sites. These findings suggest the mPBM1 protein can potentially operate as a transcriptional repressor in mouse L929 cells. Deletion mapping would allow a minimal domain responsible for this silencing effect to be delineated and aid in the isolation of mediators of this effect.

**Figure 4.4 : mPBM1 can repress transcription when tethered to a reporter gene by a heterologous DNA-binding domain in mouse L cells**

**A. The reporter and effector constructs used to assay the effect of mPBM1 on transcription**

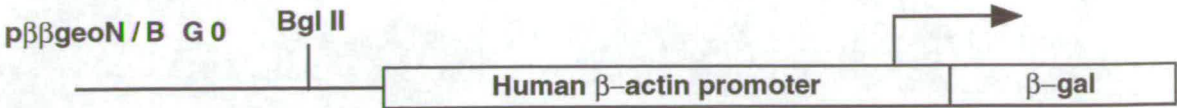
The p $\beta$ geoN/B reporter gene construct contains the human  $\beta$ -actin promoter driving expression of the  $\beta$ -galactosidase reporter gene. The p $\beta$ G5 BglII vector is identical except for five GAL4 DNA binding elements upstream of the human  $\beta$ -actin promoter. The mPBM1 MBD region was replaced with the yeast GAL4 DNA-binding domain in the pCMV-GAL4 expression construct.

**B. mPBM1 can exert a repressive effect on transcription when tethered to a promoter proximal position**

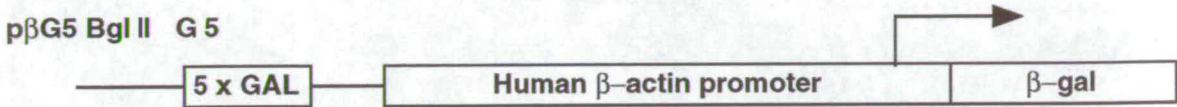
A range of concentrations (25ng-500ng) of the pCMV-GAL4-mPBM1 effector construct were co-transfected into mouse L929 cells with 2  $\mu$ g of either the p $\beta$ G5 BglII or the p $\beta$ geoN/B reporters. Expression levels of the reporter constructs are quantitated as a histogram relative to the expression of the reporter constructs in the absence of mPBM1 which is normalised to 100% for both reporters. Transcription from the p $\beta$ G5 BglII construct which contains GAL4-binding elements and is thus able to tether the GAL4-mPBM1 fusion was reduced by ~52% in the presence of 500ng of the mPBM1 fusion. In contrast no repression of transcription was observed from the p $\beta$ geoN/B vector that is unable to bind the GAL4-mPBM1 fusion.



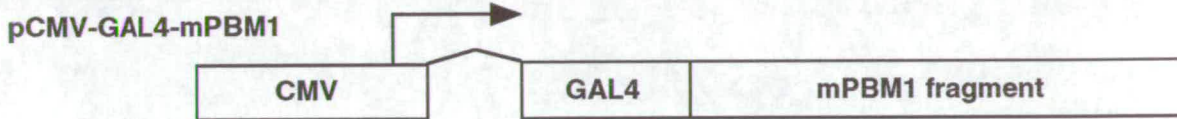
**A** Control Reporter



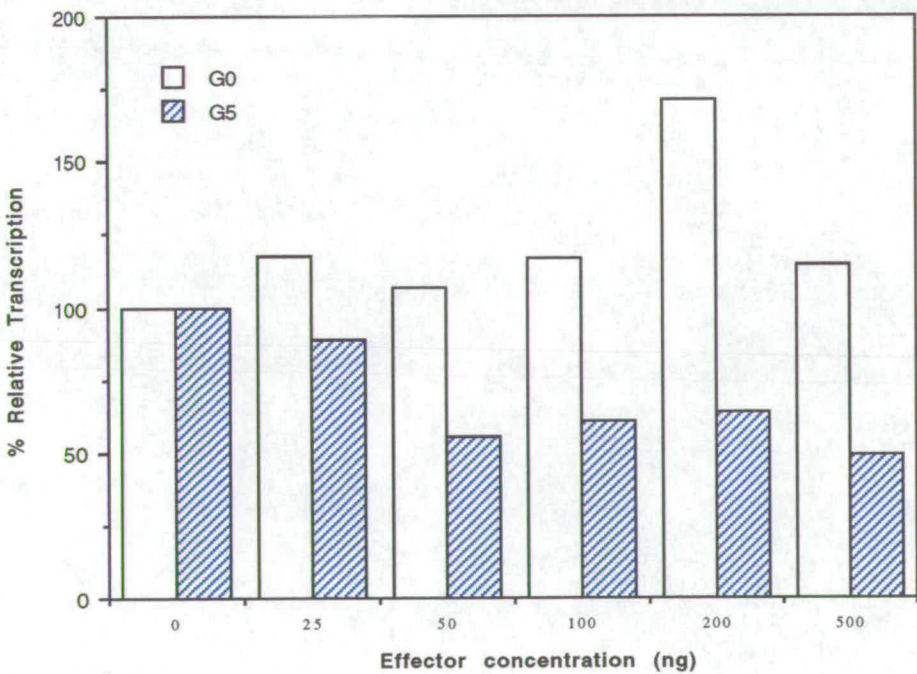
Reporter



Effector



**B**



## **4.6 Conclusions**

This section describes an investigation of the influence of the mPBM1 protein on transcription. Transfection of mouse L929 fibroblasts with increasing amounts of a GAL4-mPBM1 fusion lead to a modest dose-dependent repression of the transcriptional activity of a reporter gene containing GAL4 DNA-binding elements. Conversely, a GAL4-mPBM1 fusion exerted no repressive effect on a reporter gene lacking GAL4 binding sites. A stimulation of transcription of a reporter gene not containing GAL4 binding sites was observed following the overexpression of the mPBM1 protein. These finding are consistent with mPBM1 acting as a transcriptional repressor in mouse L929 cells.

**CHAPTER FIVE: PHYSICAL MAPPING OF THE hPBM1 GENE**

## **5.1 Introduction to mapping of human PBM1**

### **5.1.1 Chromosomal translocations involving transcriptional regulators**

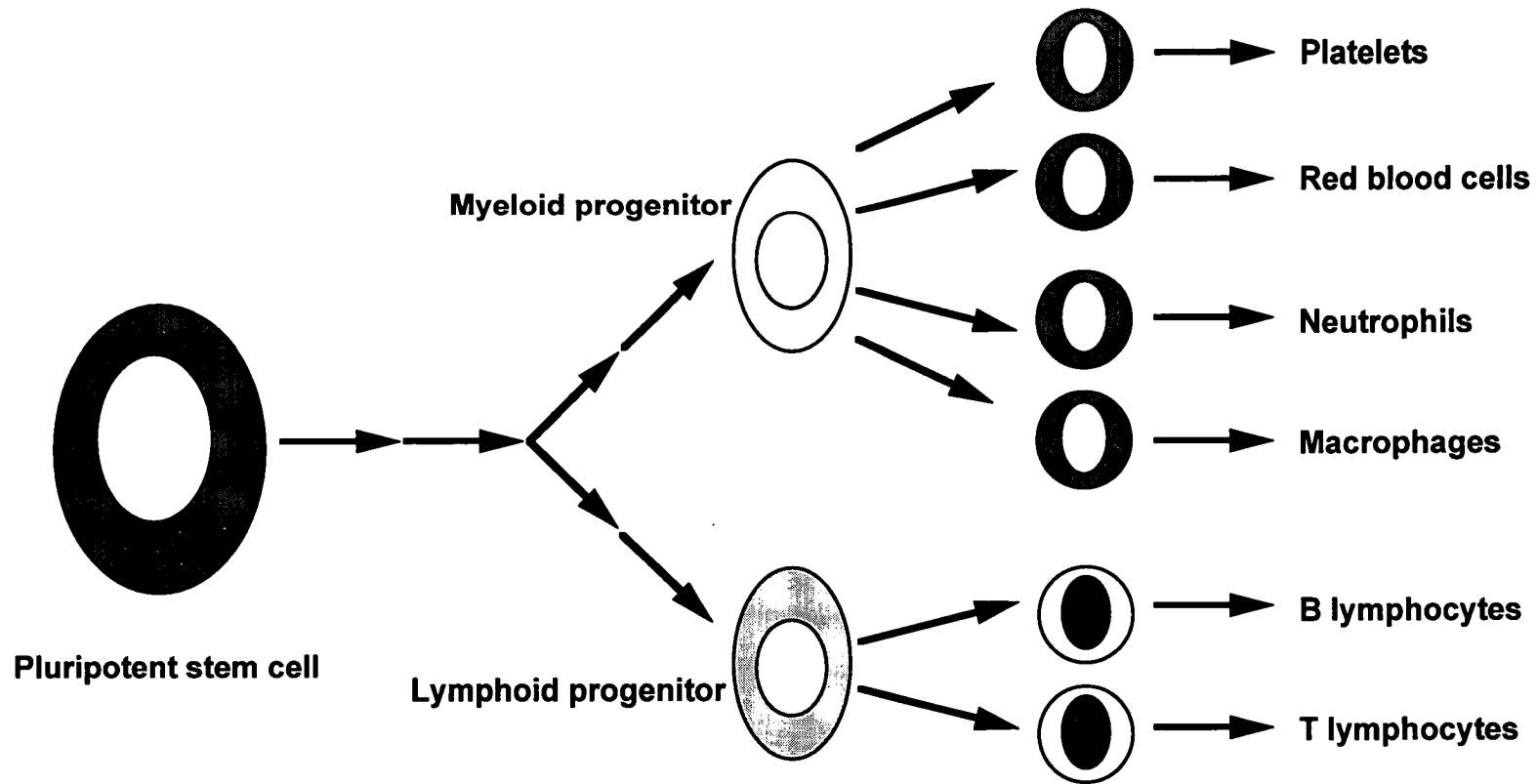
Chromosomal translocations are a consistent finding in various forms of leukaemia, lymphoma and a subset of tumours (Rabbitts, 1994). Recently it has become apparent that many of the genes that are deregulated in chromosomal translocations possess features of known transcription factors. The resultant cancers are thought to arise through the mis-expression of many downstream target genes by these gain-of-function transcriptional regulators (Cleary et al., 1992; Nichols and Nimer, 1992). This type of event can occur following breakage within intronic sequences in each of the partner genes such that the exonic sequences on either side are in-frame, resulting in a chimaeric oncofusion product. These new proteins contain domains such as various zinc fingers, AT-hooks, homeodomains and leucine zippers that appear to be critical in the development of leukaemia and other cancers.

### **5.1.2 Chromosomal translocations and acetylases/deacetylases**

Recently, Gelmetti et al., (1998) and Lutterbach et al., (1998) have reported that the leukaemia fusion partner ETO (eight-twenty-one) is fused to the acute myelogenous leukaemia 1 (AML1) transcription factor in t(8;21) AML chromosomal rearrangements. AML1 is a DNA-binding protein that can have both positive and negative effects on transcription (Ogawa et al., 1993). The t(8;21) AML1/ETO fusion protein contains the N-terminal DNA-binding domain of AML1 fused to ETO which is able to interact via a zinc finger region with the nuclear co-repressors N-CoR and SMRT and can recruit HDAC activity. Similar recruitment of N-CoR-HDAC

activity has been found in studies of retinoic acid receptor alpha (RAR $\alpha$ ) oncofusion proteins in acute promyelocytic leukaemia (APL) (Grignani et al., 1998). Thus in two separate leukaemia-signalling pathways (AML and APL), the recruitment of histone deacetylase activity by AML1/ETO and PML/RAR $\alpha$  could conceivably lead to transcriptional repression and the differentiation block at the myeloblast stage that contributes to myeloid leukaemias (see Figure 5.1).

Interestingly, the translocation t(8;16)(p11;p13) in AML involves the intrinsic histone acetyltransferase CREB-binding protein, *CBP* gene and a previously unidentified gene termed *MOZ* (Borrow et al., 1996). CBP and the closely related p300 transcriptional coactivator protein are histone acetyltransferases (Ogryzko et al., 1996) while MOZ possesses a putative acetyltransferase domain. The position of the translocation breakpoint within CBP maintains the CREB-binding domain and an N-terminal region necessary for transactivation (Chrivia et al., 1993) and fuses in-frame to MOZ maintaining the putative acetyltransferase signature motif and nuclear localisation signals. These findings suggest that the CBP-MOZ oncofusion product would result in leukaemogenesis through aberrant chromatin acetylation at promoters requiring CBP leading to a loss of control over normal cellular growth and differentiation potential.



**Figure 5.1 :** The cell lineages of the haematopoietic system

All haematopoietic lineages derive from a pluripotent stem cell through lineage-committed intermediates. Chromosomal translocations that perturb normal haematopoietic development lead to leukaemias which are uncontrolled proliferations of haematopoietic cells that do not retain the capacity to differentiate to mature blood cells.

### **5.1.3 Chromosomal translocations at 11q23 - the ALL1 locus**

Curiously, many fusion proteins arising from chromosomal translocations show a specificity between the target genes, however some loci appear to rearrange promiscuously with numerous partner loci. For example, the ALL1 locus at 11q23 rearranges with as many as 25 different chromosomal regions generating a wide range of abnormalities in approximately 5% of acute leukaemias (Rabbitts, 1994). Chromosomal translocations involving band 11q23 are the most common cytogenetic abnormality in children with acute leukaemia. The ALL1 protein has numerous motifs homologous to proteins implicated in transcriptional regulation such as an N-terminal AT-hook, a cysteine rich region homologous to the Dnmt1 methyltransferase, an extreme C-terminal SET domain and a centrally located cysteine rich region homologous to *Drosophila* trithorax that has been variously called a plant homeodomain (PHD)-zinc finger, trithorax consensus (TTC) and the leukaemia-associated protein (LAP) zinc finger (Aasland et al., 1995; Koken et al., 1995; Saha et al., 1995).

### **5.1.4 Chromosomal rearrangements of ALL1 - the PHD-zinc finger**

When Djabali et al., (1992) identified the ALL1 gene at 11q23, it was apparent that the zinc finger motifs did not correspond to the zinc binding domains found in genes previously implicated in chromosomal translocations, such as the LIM, RING and Kruppel zinc finger domains (Sanchez-Garcia and Rabbitts, 1994) and was in fact the widely conserved PHD-finger motif. Intriguingly this domain was found in tandem in 11q23 partner genes on chromosome 10 (AF10) and chromosome 17 (AF17) in the



translocation t(10;11)(p12;23) (Chaplin et al., 1995) and t(11;17)(q23;q21) (Prasad et al., 1994) rearrangements respectively. The basic domain consists of a C<sub>4</sub>HC<sub>3</sub> zinc finger motif (C-X<sub>2</sub>-C-X<sub>n</sub>-C-X<sub>2</sub>-C-X<sub>4</sub>-H-X<sub>2</sub>-C-X<sub>n</sub>-C-X<sub>2</sub>-C) that has been proposed to fold into a single domain that is stabilised by two zinc atoms and is found in a diverse array of proteins conserved from yeast to humans. In all t(10;11) and t(11;17) translocations documented, the derivative ALL1-AF10 and ALL1-AF17 fusions encode proteins completely lacking PHD-zinc finger domains.

The aberrant recruitment of either acetyltransferase or deacetylase activities following chromosomal rearrangements has been a frequent finding, suggesting that normal patterns of histone acetylation might be perturbed by the resultant oncofusion products leading to leukaemia. The PHD-zinc finger of Mi2 $\beta$  does associate with HDAC activity and requires an intact zinc finger domain to do so (Zhang et al., 1998) and the oncofusion products from 11q23 rearrangements with various target loci are frequently found to lack this motif suggesting a critical role for this conserved zinc finger domain in myeloproliferation.

Thus, a wealth of evidence exists suggesting the PHD-zinc finger domain may be involved in leukaemogenesis. It was of interest, given the presence of a PHD-zinc finger domain at the C-terminus of mPBM1, to localise chromosomally the human PBM1 gene to determine if rearrangements or deletions were found in cytogenetic abnormalities leading to leukaemia.

## 5.2 Results

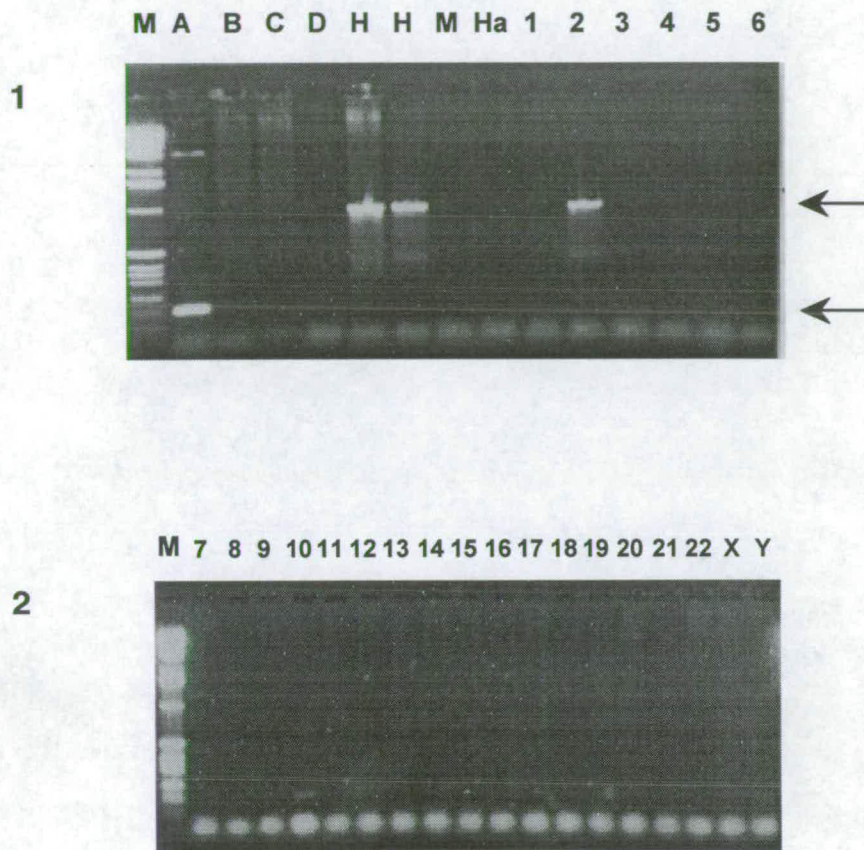
### 5.2.1 Mapping of hPBM1 using monochromosomal cell hybrids

Using the UniGene public server (<http://www.ncbi.nlm.nih.gov/Schuler/UniGene>), a cluster of human ESTs (593380 and 561792) were identified that were highly homologous to mPBM1 and were importantly distinct from hPBM2 (KIAA0314). Specific primers to the human EST 593380 that encodes an MBD highly homologous to mPBM1 (see Figure 3.13) were designed to amplify a 100bp product from cDNA.

593380F : 5'-ATGAACGTGAACTGCGTATTCC-3'

593380R : 5'-TTTATTACTTCAGGGTACTGCC-3'

Figure 5.2 shows that these primers amplified a 142bp fragment from a cDNA clone but amplified a ~0.8kb product from human genomic DNA. This fragment was sub-cloned and sequenced and the position of an internal intron within the MBD assigned (see Figure 3.13). A monochromosomal human/rodent cell hybrid panel of DNAs (obtained from the UK-HGMP Resource Centre, Cambridge, UK) was screened with the 593380F and 593380R primers. To avoid cross hybridisation and non-specific PCR amplification from related genes, forward primer with template, reverse primer with template, both primers without template and mouse, hamster and human DNAs with both primers were amplified in parallel with the genomic DNAs from the monochromosomal cell hybrids. A product was amplified from human genomic DNA and DNA from a hybrid containing human chromosome 2 (see Figure 5.2).



**Figure 5.2 : Mapping of the human PBM1 gene to chromosome 2**

Chromosomal assignment of the hPBM1 gene was performed by PCR analysis of human/rodent somatic cell hybrids using primers 593380F and 593380R that amplify a 142bp product from cDNA and a 833bp fragment from genomic DNA. In panel 1, the 593380 primers amplified a ~0.8kb product from human genomic DNA and DNA derived from a hybrid containing human chromosome 2. No product was amplified from mouse or hamster DNA or DNA from hybrids containing the remaining human autosome and sex chromosomes.

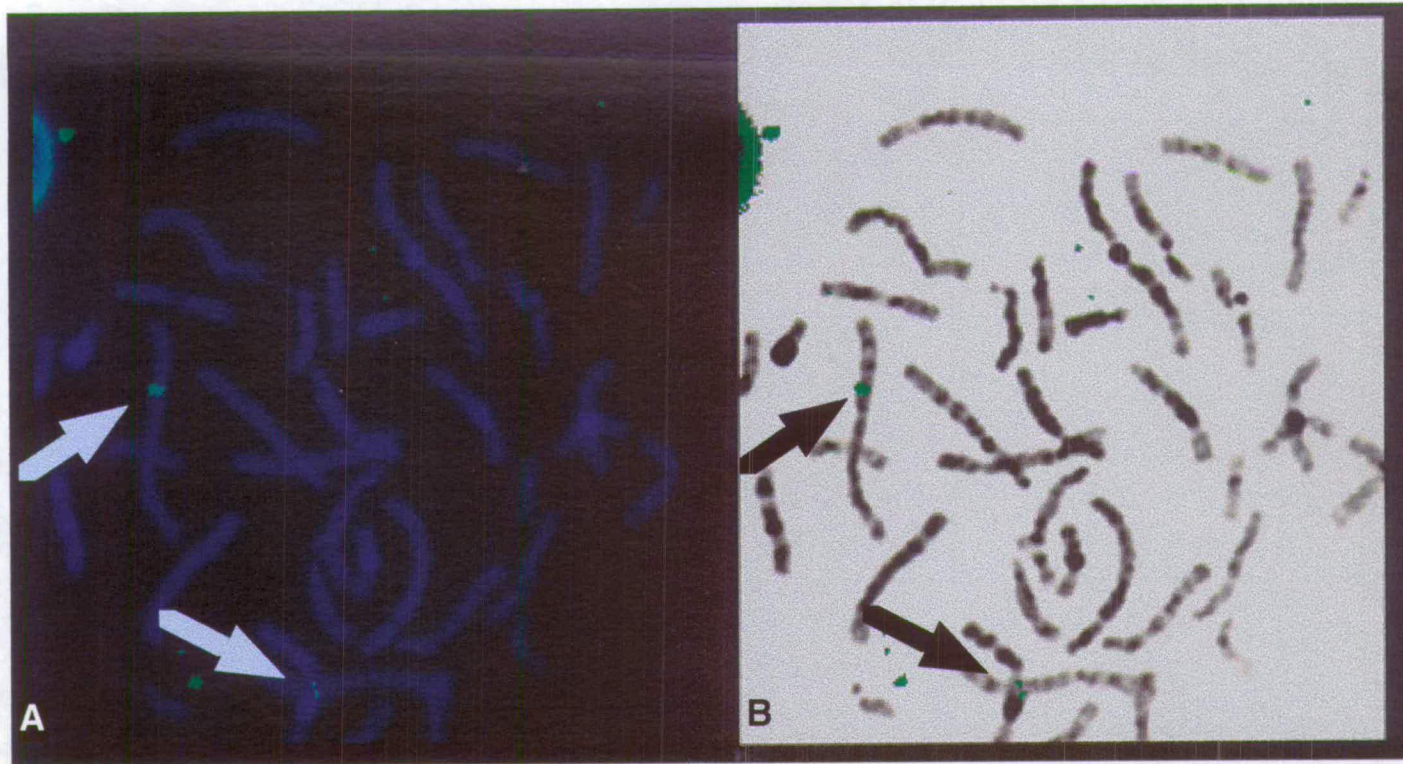
### 5.2.2 Mapping of hPBM1 using fluorescence *in situ* hybridisation (FISH)

To confirm the chromosomal localisation of the hPBM1 gene on chromosome 2 and further refine the map position, a specific genomic clone was isolated for fluorescence *in situ* hybridisation (FISH). A human P1 artificial chromosome (PAC) genomic library (obtained from UK-HGMP and constructed by Pieter de Jong, Roswell Park Cancer Institute, Buffalo, USA) was screened with primer pairs 593380S / 593380AS and 561792F / 561792R.

593380S	:	5'-AGCTGGCAGAGAGAGACAAG-3'
593380AS	:	5'-CTTTCCACATGGAGCATAAT-3'
561792F	:	5'-ACACTCTTGCTGCATGCCAGTTC-3'
561792R	:	5'-ACTGAAAGTCTGAAGACCAAAG-3'

A PAC clone, P22 was identified by screening the PAC library using the 561792 primer pairs and the presence of the hPBM1 gene within the P22 PAC clone was confirmed by sub-cloning and sequencing. The hPBM1 P22 PAC clone was labelled for probing by incorporation of biotin-16-dUTP. Following incubation of the biotin labelled probe on human metaphase spreads, the PAC clone was detected by avidin-FITC and the signal was amplified by using a biotinylated anti-avidin antibody and a second incubation with avidin-FITC. FISH clearly showed a unique hybridisation signal on numerous metaphase spreads corresponding to the human G-band at chromosome 2q23 (Figure 5.3). Thus by two independent mapping approaches the human PBM1 gene maps to chromosome 2.



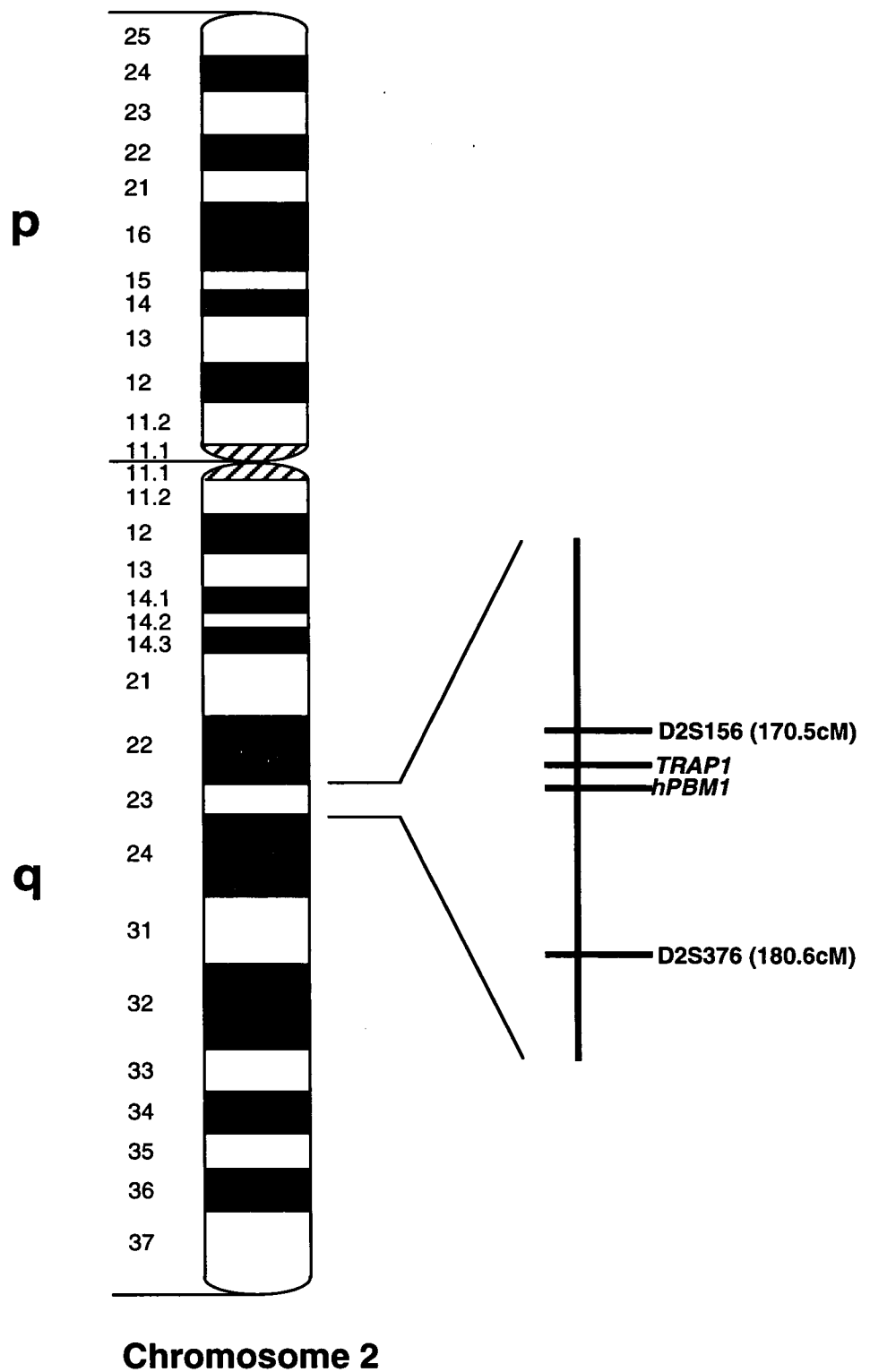


**Figure 5.3 :** Fluorescence *in situ* hybridisation with PAC P22, containing the hPBM1 gene, on human metaphase spreads

A. PAC P22 labeled with biotin-16-dUTP and detected by avidin-FITC. The signal was amplified by using a biotinylated anti-avidin antibody and a second incubation with avidin-FITC. Chromosomes are stained with DAPI.

B. An inverted DAPI banding pattern showing the G-banding pattern of the human karyotype. Arrows show a signal on human chromosome 2q23.

The mapping of human PBM1 was carried out in collaboration with Shelia Boyle (MRC-Human Genetics Unit, Edinburgh).



**Figure 5.4 :** The human PBM1 gene maps to chromosome 2q23

Chromosomal placement of the human PBM1 gene in a 10cM genetic interval between the microsatellite markers D2S156 and D2S376. The approximate corresponding location of the hPBM1 gene within the q23 region is indicated relative to a cytogenetic ideogram of human chromosome 2.

### **5.2.3 Mapping of hPBM1 by radiation hybrid mapping**

Recently, the Whitehead Institute/MIT Centre for Genome Research radiation hybrid mapping project has independently confirmed the assignment of the ESTs 593380 and 561792 derived from the hPBM1 locus to chromosome 2. Using a sequence tagged site (STS) approach utilising radiation hybrid mapping, ESTs 593380 and 561792 were placed in a 10 cM interval between the microsatellite markers D2S156 (distal; 170.5 cM) and D2S376 (proximal; 180.6 cM) on chromosome 2 downstream of the human TRAF-interacting protein 1 (TRAP1) gene that is within a region co-linear with 2q23 (see Figure 5.4).

### **5.3 Conclusions**

The hPBM1 gene was mapped by monochromosomal hybrid PCR to human chromosome 2. The location of hPBM1 was refined by the isolation of a PAC clone and FISH analysis which placed the hPBM1 gene on the long arm of chromosome 2 at 2q23. An STS approach by the Whitehead Institute/MIT Centre for Genome Research radiation hybrid mapping project has independently confirmed the assignment of ESTs 593380 and 561792 from hPBM1 to a 10 cM interval between the microsatellite markers D2S156 (distal; 170.5 cM) and D2S376 (proximal; 180.6 cM) on chromosome 2 downstream of the human TRAF-interacting protein 1 (TRAP1) gene in a region co-linear with 2q23. The mapping of hPBM1 to 2q23 makes it a candidate for disruption in the cases of leukaemia involving rearrangements and deletions on the distal arm of chromosome 2 (See discussion in Chapter 7).

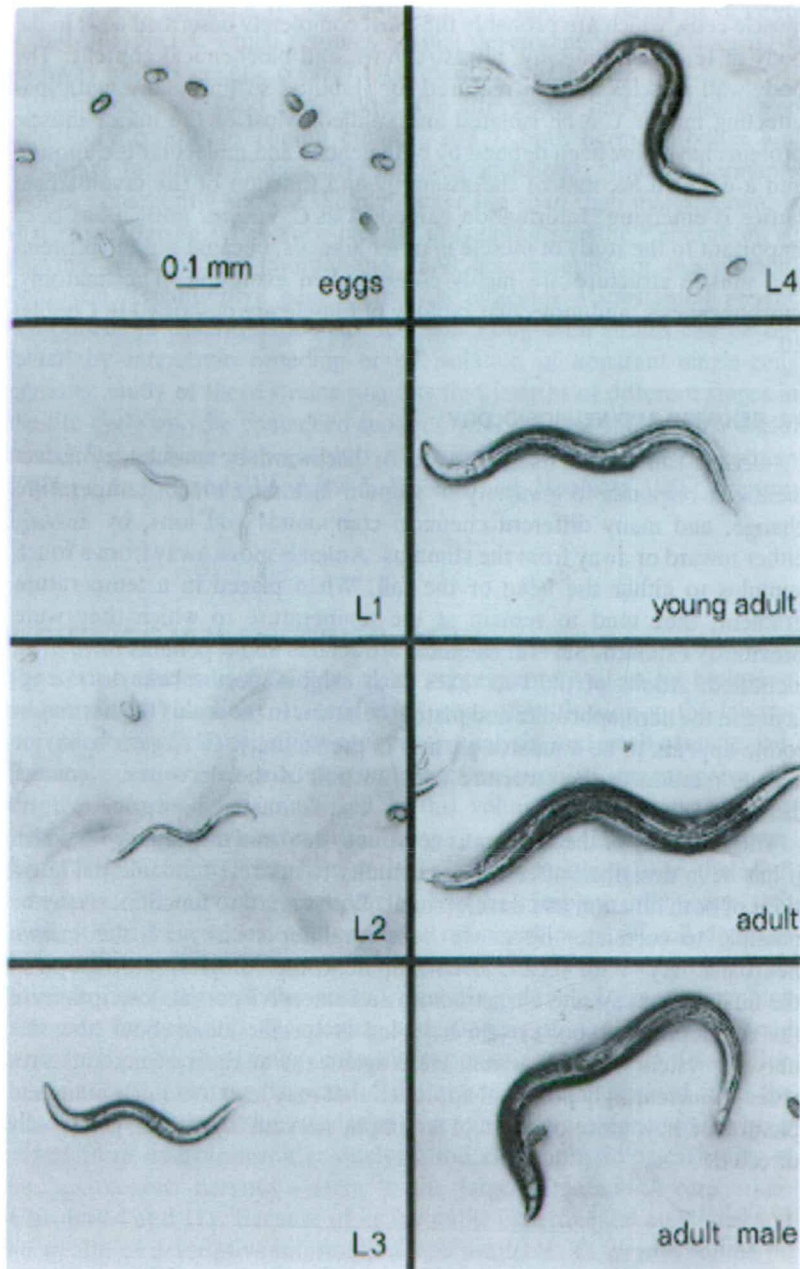


**CHAPTER SIX: ISOLATION OF A *pbm-1* NULL MUTANT**

## 6.1 Introduction

*Caneorhabditis elegans* is a small, free-living soil nematode that is distributed worldwide. The propagation of the animal does not require mating as *C. elegans* is a self-fertilising hermaphrodite producing both sperm and oocytes. Males do however arise spontaneously at a low frequency (<0.2% males from wild type hermaphrodites) as a result of meiotic non-disjunction (Hodgkin et al., 1979) and are particularly useful for genetic crosses. Hermaphrodites and males are approximately 1mm long and males are readily distinguishable on the basis of behaviour and morphology, with males possessing a distinctive tail which has a fan-shaped structure as opposed to the elongated tail spike of the hermaphrodite. An adult hermaphrodite worm lays around 300 eggs during its reproductive life. Worm larvae hatch from eggs and proceed through a series of four larval stages, L1 to L4 and the adult (L4) hermaphrodite worms are fertile (see Figure 6.1).

*C. elegans* was chosen originally as a model system to completely describe the development of a simple metazoan (Brenner, 1974). That is to describe the molecular mechanisms that maintain the determined and differentiated states of cellular lineages from the incipient embryo to the adult animal. Sulston et al., (1983) subsequently presented a complete description of the entire embryonic lineage from the zygote to the hatched larva. Briefly, following fertilisation of an oocyte by a sperm, the zygote develops by a series of asymmetric divisions that establishes the division of somatic and germ cells in the developing embryo. The first cleavage of the zygote produces a



**Figure 6.1** : Eggs, larvae and adult *Caenorhabditis elegans*

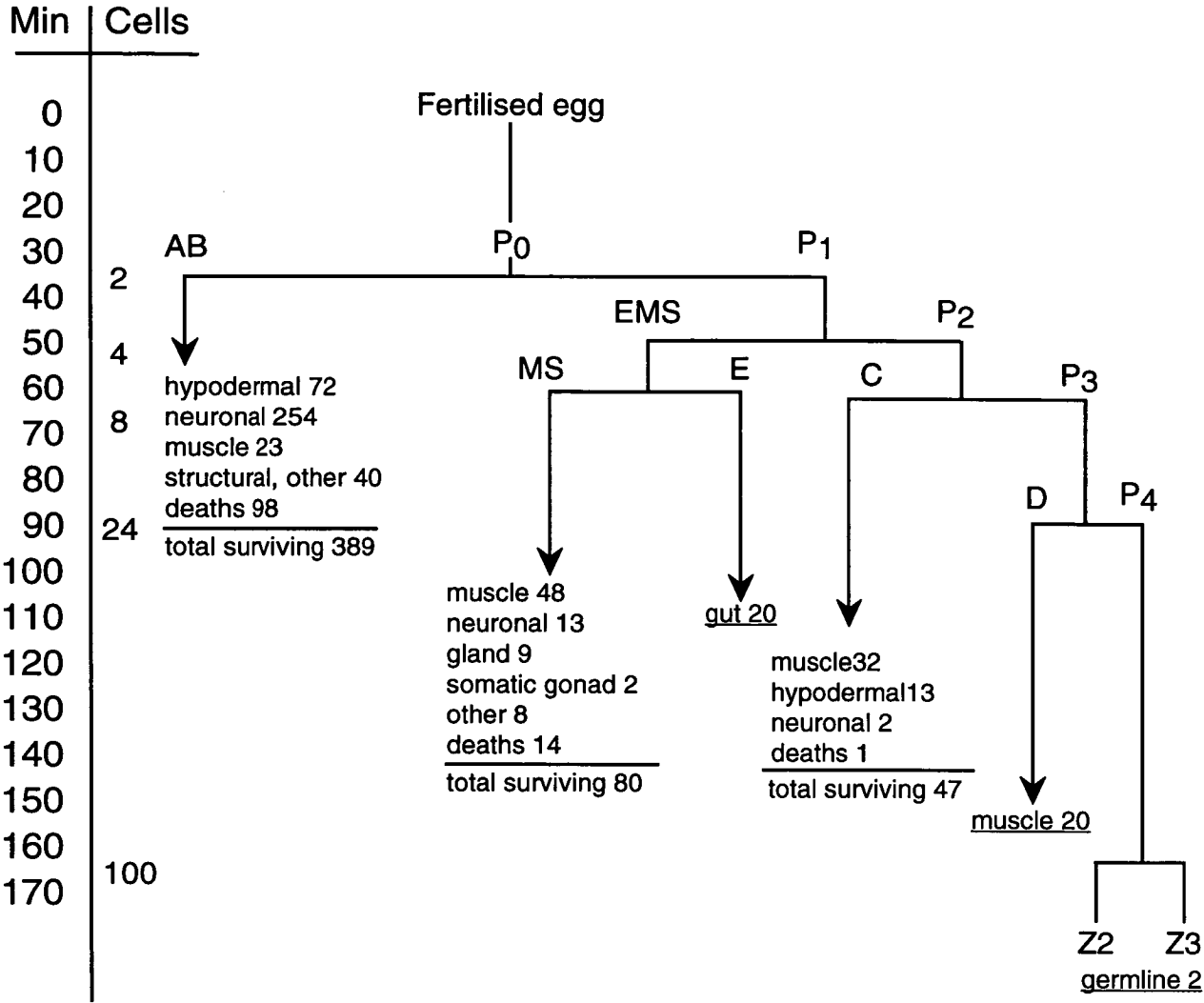
The larval stages, L1 - L4 of the hermaphrodite *C. elegans* are depicted. An adult male is shown in the lower right panel with the characteristic copulatory ray on the tail.

This Figure was taken from Wood (1988).

large anterior founder cell, AB and a smaller posterior founder cell P1. This P1 cell gives rise ultimately to four somatic founder cells MS, E, C and D and the germline precursor P4 and with AB, these blastomeres give rise to all the adult tissues of the worm (see Figure 6.2). Thus, the cell types derived from these initial founder cells and a complete description of the cell lineages has been traced from the zygote to the 959 somatic cells that comprise the adult hermaphrodite worm. In addition, a complete description of the wiring of the nematode nervous system of the 302 neuronal cells exists down to the synaptic level (White et al., 1987).

*C. elegans* is the first multicellular organism for which an essentially complete genome sequence has been determined (The *C. elegans* sequencing consortium, 1998). The total 97-megabase genome is approximately one-thirtieth the size of the human genome and eight times the size of the genome of the budding yeast, *Saccharomyces cerevisiae*. The *C. elegans* genome contains 19 099 genes predicted to encode a functional protein product. Some 42% of these protein products contain distant matches outside the nematode lineage (The *C. elegans* sequencing consortium, 1998).

More than 70 human disease genes have been positionally cloned and the majority contain statistically significant sequence identity to genes in model organisms (Mushegian et al., 1997). Thus, while many newly identified mammalian sequences possess homologous sequences in the model organisms it is necessary to distinguish whether these homologs in distant organisms represent direct counterparts with high



**Figure 6.2 :** Early lineage pattern of cleavages in the *C. elegans* embryo

A series of asymmetric cleavages of the developing embryo generates the six blastomeres : AB, MS, E, C, D and P<sub>4</sub> from which the differentiated tissues of the adult animal derives (Sulston et al., 1983).

This Figure was taken from Wood (1988).

identity across the sequence and the same modular organisation of motifs (a true ortholog) or are the sequences simply homologous with the similarity restricted to certain isolated regions of the sequence and certain domains. Strikingly, it has been found that the nematode worm, *C. elegans* encoded true orthologs in 36% of cases (Mushegian et al., 1997) and it has been estimated that 50% of human genes have some homology to a *C. elegans* gene (Ahringer, 1997). The assignment of human disease genes with orthologous genes in model organisms thus represents a comparative genomic approach where function can be elucidated by exploiting the forward and reverse genetic strategies possible in *C. elegans*, *Drosophila* and the yeasts. As a result, a systematic multi-organismal approach has been initiated to cross-reference sequences in model organisms with mammalian cDNAs to accelerate functional analysis (Bassett et al., 1997).

In order to study the function of *pbm-1*, a mutation in the gene was required. There are three main methods to analyse gene function in the nematode by inhibiting its activity : 1) RNA interference (RNAi), 2) identification of a transposon insertion mutant and the subsequent isolation of a strain harbouring a deletion and 3) isolation of a chemically induced deletion mutant directly.

Fire and co-workers (1998) have demonstrated that microinjection of double-stranded (ds) RNA into the worm syncytial germline leads to a potent and specific interference of gene activity. This RNAi technique has been demonstrated to phenocopy

previously identified loss of function or hypomorphic alleles in the targeted gene. However, despite the usefulness of RNAi, particularly in the study of multigene families, the technique suffers from the caveat that only the injected animal and the F<sub>1</sub> progeny exhibit the effects of RNAi whereas the F<sub>3</sub> progeny typically revert to the wildtype phenotype (Fire et al., 1998). The RNAi technique appears to operate at a post-transcriptional level (Montgomery et al., 1998) and as a result the effect is non-heritable. For genetic and functional analysis of *pbm-1* a mutant was required.

To obtain a *pbm-1* mutant, the reverse genetic strategy of target-selected gene mutagenesis was employed to 1) identify worm strains containing either transposon element (Tc1) insertions within the *pbm-1* gene from which a deletion strain could be derived (Zwaal et al., 1993) and 2) the direct isolation of a worm strain containing a deletion within *pbm-1* induced by chemical mutagens (Jansen et al., 1997). These reverse genetic procedures for target-selected gene inactivation take advantage of the fact that *C. elegans* can survive frozen storage (Wood et al., 1988). As a result large scale mutagenisation protocols can be devised where following the identification of mutants by PCR, the siblings of the animal from which the genomic template was derived can be recovered for propagation for functional analysis.

Tc1 transposon mutagenesis in *C. elegans* is derived from the use of P element insertion in *Drosophila* (Ballinger and Benzer, 1989). This method takes advantage of the genome of *C. elegans* containing numerous copies of the Tc1 transposon which is



mobile in certain mutator strains. Screening the Tc1 insertion library is by PCR to detect Tc1 insertions within a gene of interest. Isolation of a strain from which the DNA template used in the amplification reaction was derived and the subsequent mobilisation of the element with removal of flanking DNA generates a deletion within the gene of interest (Zwaal et al., 1993) (see Figure 6.3).

## **6.2 Results**

### **6.2.1 Identification of Tc1 transposon insertions in *pbm-1***

The Tc1 insertion mutant library consists of three sets of 960 cultures of MT3126 (*mut-2*) worms and an equivalent set of genomic DNAs arranged as an ordered matrix of 10 racks containing 12 X 8 trays. An ordered pooling of aliquots of the above genomic DNAs reduced these 960 lysates to 8 pools of 120 (see Figure 6.4A). The screening strategy to identify Tc1 insertions within the *pbm-1* gene employed PCR using a primer specific to the inverted repeats of the transposon and a *pbm-1* specific primer. Nested internal primers specific to Tc1 and *pbm-1* were used to improve the specificity and the sensitivity of the screen. Following the amplification of the intervening DNA, the reactions were loaded onto agarose gels to visualise Tc1 insertion events. Two sets of nested primers were designed to be 3.0kb apart at the 5' and 3' ends of the *pbm-1* locus, to 1) maximise coverage of the screen to detect Tc1 insertions and 2) such that following the identification of insertions, these primers could be used to detect any deletions arising from excision of the transposon.

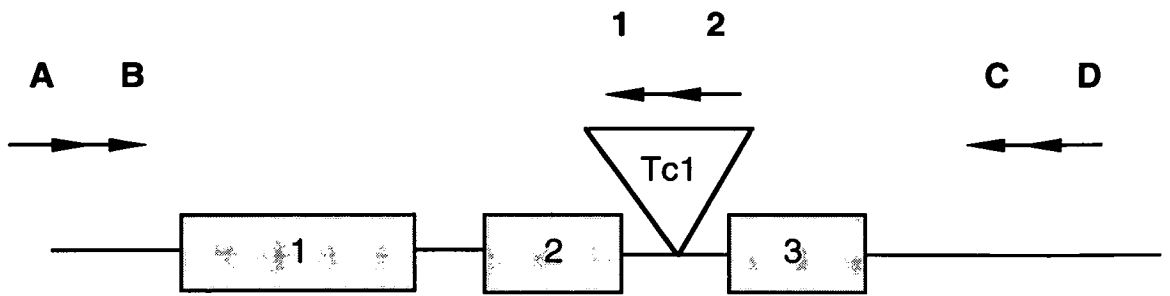
**Figure 6.3 : Reverse genetic strategy for target-selected gene inactivation using transposon mutagenesis**

Transposon insertion events within a gene of interest are identified by a nested PCR-based approach. Boxes represent exons, lines intervening introns in a hypothetical three exon gene and triangles a Tc1 transposon insertion event. A Tc1 insertion mutant library consists of sets of polyclonal cultures of MT3126 (*mut-2*) worms and equivalent pools of genomic DNA derived from these frozen stocks.

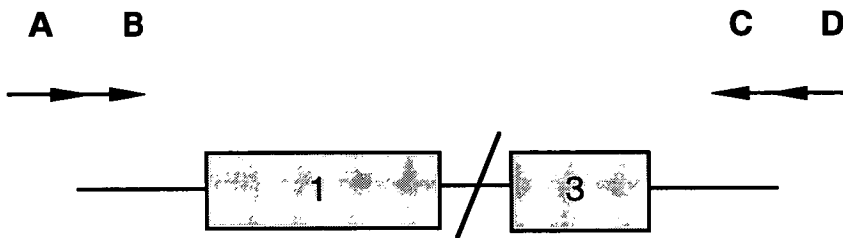
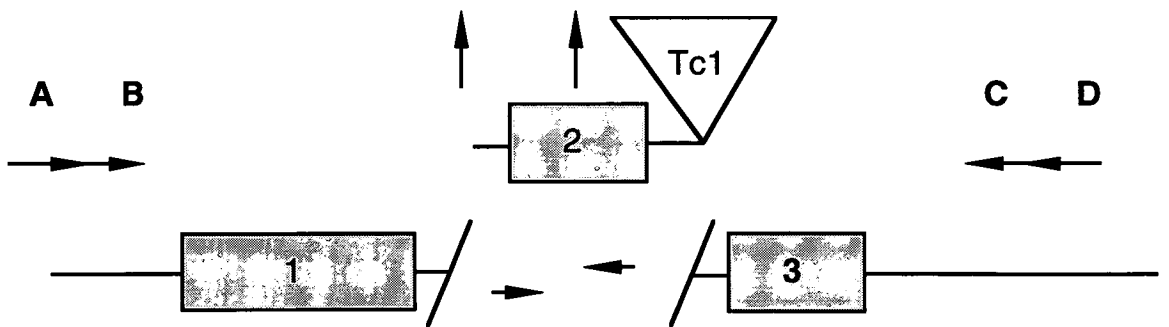
1. Genomic DNA from a transposon mutant bank is screened with gene-specific (A, B, C and D) and transposon (Tc1)-specific primers (1 and 2). Consistent amplification of a DNA product by PCR identifies a corresponding frozen stock of nematodes in an ordered matrix. This oligoclonal population of nematodes are thawed and seeded individually to NGM plates. These worms are screened by PCR and a strain harbouring a Tc1 insertion within the gene of interest identified.

2. F<sub>1</sub> progeny from these animals are tested by PCR for transposon excision events and the removal of flanking DNA by amplification across the site of the transposon insertion such that the deletion product has a selective advantage in the reaction. Sibling selection is carried out to maintain strains identified. The mechanism of Tc1 transposon-induced deletion is not known but is thought to arise as a result of repair of the double strand break following transposon excision.

1



2



### Tc1 specific primers

L1 : 5'-CGTGGGTATTCCTTGTTCTGAAGCCAGCTAC-3'  
L2 : 5'-TCAAGTCAAATGGATGCTTGAG-3'  
R1 : 5'-TCACAAGCTGATCGACTCGATGCCACGTCG-3'  
R2 : 5'-GATTTTGTGAACACTGTGGTGAAG-3'

### *pbm-1* specific primers

ZK1 : 5'-TAATCAGTTTCTACTTCTGC-3'  
ZK2 : 5'-GCAACAACAACCTTCTACAGC-3'  
ZK3 : 5'-CACCATCCTCTCTTCATATG-3'  
ZK4 : 5'-AGAATCCCGCTGCACTGACC-3'

783.4-1 : 5'-AGATGGAGATGCTAATCACG-3'  
783.4-2 : 5'-AGGCAACTATTAAGACTAGC-3'  
783.4-3 : 5'-TGATCAATGTTGAATTGACG-3'  
783.4-4 : 5'-GTTGTATTTTCAGCTGCTTCC-3'

The screen was carried out in quadruplicate to distinguish Tc1 insertions within *pbm-1* from spurious amplification products. A positive address shows four identically sized products and four positive addresses G I, G II, F III and H II are indicated in Figure 6.4C corresponding to Tc1 insertion events within the *pbm-1* gene.

In this instance, “positive address” refers to the reproducible amplification in quadruplicate of a PCR fragment corresponding to DNA derived from the transposon and the *pbm-1* gene from genomic DNAs that correspond to a frozen stock of polyclonal nematodes. This can be confirmed directly by sequencing and is a useful way to determine the nature of the Tc1 insertion event (i.e. intronic or exonic). Following the identification of a positive address, DNA lysates of lower complexity are animals until eventually a frozen stock of oligoclonal nematodes are thawed.

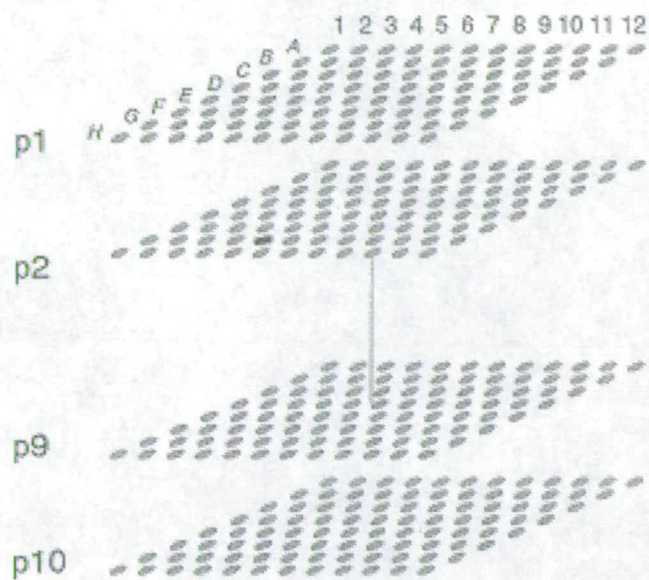
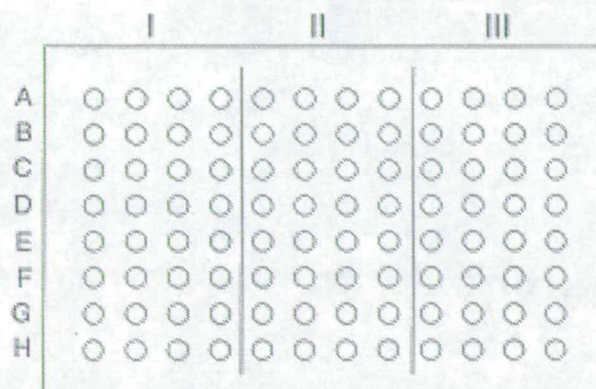
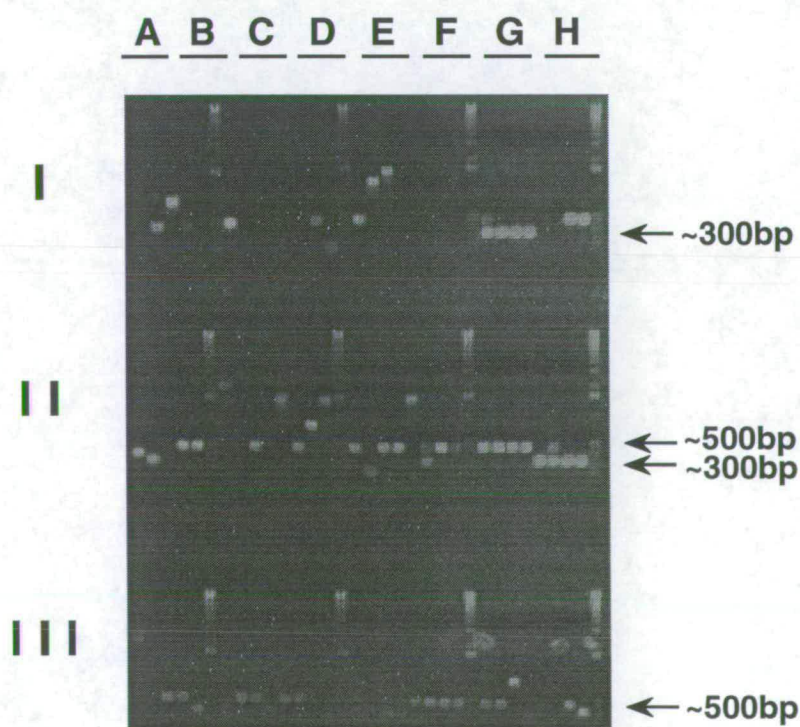
**Figure 6.4 : Identification of Tc1 insertion events within the *pbm-1* gene**

A. The Tc1 insertion mutant library consists of three sets (I, II and III) of 960 cultures of MT3126 (*mut-2*) worms and an equivalent set of genomic DNAs arranged as an ordered matrix of 10 racks containing 12 X 8 trays. An ordered pooling of aliquots of the above genomic DNAs reduced these 960 lysates to 8 pools (A-H) of 120.

B. A location within this matrix ("an address") is identified by PCR screening of these DNA pools using primers specific to the inverted repeats of the Tc1 transposon and to the gene of interest.

C. These 8 pools of 120 were screened with the *pbm-1* specific primers (ZK1 to 4 and 783.1-1 to 4) and the Tc1-specific primers (L1, L2 and R1, R2) in all possible combinations to identify transposon insertion events within *pbm-1*.

The primers employed in the gel shown are *pbm-1* specific primer 783.4-4 and Tc1 transposon specific primer R1 then nested PCR with primer pairs 783.4-3 and R2 (see p178 for primer sequences). The screen was carried out in quadruplicate to distinguish Tc1 insertions within *pbm-1* from spurious PCR products. A positive address shows three or four identically sized products. Four positive addresses G I, G II, F III and H II are indicated which correspond to Tc1 insertion events within the *pbm-1* gene.

**A****B****C**

Individual animals are picked to seeded NGM plates and the strain containing a Tc1 insertion in *pbm-1* is directly identified. The progeny from this Tc1 insertion mutant should contain deletion derivatives in the MT3126 (*mut-2*) strain arising from mobilisation of the element and removal of flanking DNA. These deletion events can be visualised by PCR across the site of insertion (Zwaal et al., 1993).

Thus, isolation of a deletion in *pbm-1* from the Tc1 transposon insertion bank is a two step procedure requiring first, the identification of a strain containing a Tc1 insertion event within the gene and second, the isolation of a deletion derivative from this strain. Moreover, Tc1 transposons are known to preferentially intergrade within the T + A-rich introns in *C. elegans* (Rosenzweig et al., 1983; Zwaal et al., 1993). This non-random insertion bias of Tc1 transposons within intronic sequences therefore does not guarantee isolation of a loss-of function mutant. The identification of a *pbm-1* loss-of-function insertion mutant was discontinued at this stage, for the above reasons, and because a chemical deletion mutant was identified in parallel with the screening of the transposon insertion library.

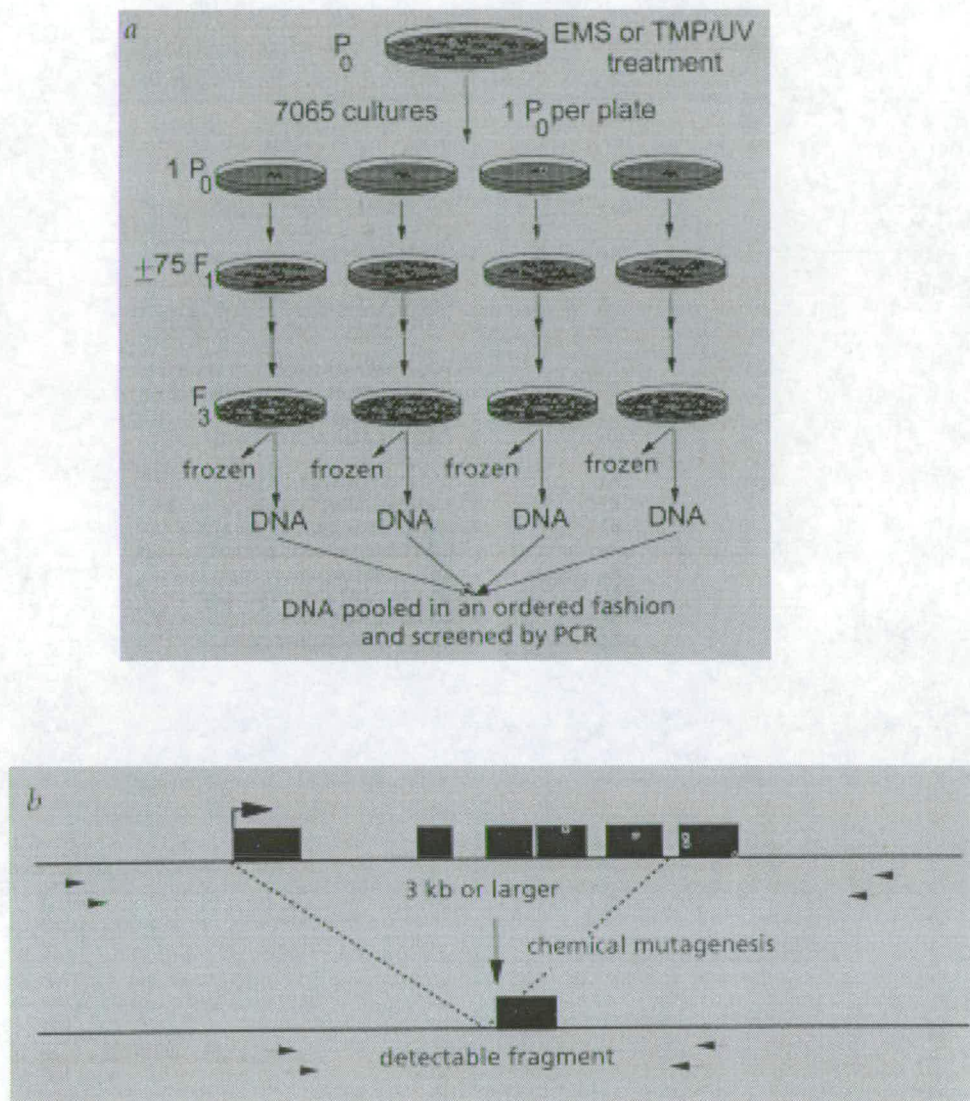
The deletion mutant library possesses the distinct advantage that it allows the direct isolation of a deletion within the *pbm-1* gene in a single step. Furthermore, the mutagenesis procedure is random generating deletions in a non-specific manner (Jansen et al., 1997). This deletion mutant approach was chosen to isolate a *pbm-1* loss of function mutant.



### 6.2.2 Identification and isolation of a *pbm-1* deletion mutant

Plasterk and co-workers (Jansen et al., 1997) have described an improved method for target-selected gene inactivation using chemical mutagens. This procedure involves the random mutagenesis of worms with ethyl methanesulphonate (EMS) or trimethylpsoralen (TMP) both of which can generate deletion mutations (Anderson, 1995; Yandell et al., 1994). Mutagenised populations of animals are propagated until the F<sub>3</sub> generation, then divided for freezing as a stock and for the preparation of a corresponding DNA lysate. These genomic DNA lysates derived from the mutagenised worms are arranged in an ordered matrix and screened by PCR to detect deletions (see Figure 6.5). The term “positive address” here refers to the reproducible amplification of a deletion fragment smaller than the wild type from genomic DNAs derived from a frozen stock of polyclonal nematodes. Each address contains multiple independent mutagenised genomes (Jansen et al., 1997). The deletion fragments importantly possess a selective advantage in the PCR compared to the larger wild type fragments. As the screen is by PCR on genomic DNA a round of sibling selection is carried out to maintain corresponding worm stocks. These DNA lysates were screened with the *pbm-1* primers presented in section 6.2.1.

Figure 6.6A shows the PCR screening of the deletion mutant library in quadruplicate with nested *pbm-1* specific primers (ZK1, ZK2, ZK3 and ZK4) that amplified a wild type fragment of 2961bp from the 5' end of the gene. In most lanes the amplification of a wild type fragment of 2.9kb is seen. The failure of the ZK1-4 primers to amplify



**Figure 6.5 : Reverse genetics by chemical mutagenesis**

A. Synchronized worms are treated with either 50mM EMS or 30 $\mu$ g/ml trimethylpsoralen and then transferred to NGM plates. F3 progeny are washed off these plates and divided in three, twice for freezing polyclonal nematode stocks and once for isolation of genomic DNA for library construction. These crude lysates are pooled in a ordered matrix for PCR screening.

B. Deletions in a hypothetical gene of interest are screened by nested PCR across the gene locus using primers that amplify a 3kb fragment.

This figure was taken directly from Jansen et al., (1997).



a wild type fragment in all cases is likely due to the quality of the genomic DNA template that degrades as the number of freeze-thaw cycles increases (Karen Thijssen, personal communication). The screen is carried out 4-fold to distinguish deletion fragments from artefacts of the PCR amplification procedure. In lanes 5H (3/4 positive deletion fragments), 8H (2/4) and 10B (2/4) the *pbm-1* primers amplified a deletion fragment of approximately 1.0kb. This suggested that the corresponding stock of frozen worms from which the genomic DNA was derived contained an animal(s) with a large (~2.0kb) 5' deletion in the *pbm-1* gene. The *pbm-1* primers specific to the 3' end of the gene (783.4-1, 783.4-2, 783.4-3 and 783.4-4) amplified a wild type fragment of ~3.0kb in the majority of cases but no deletion fragments corresponding to deletions at the 3' end of the gene were identified (data not shown). The addresses 5H and 10B were found to reproducibly amplify a ~1.0kb deletion fragment and were analysed further. No additional deletion fragments that could conceivably represent further deletion alleles of *pbm-1* were detected in this screen of the mutant bank.

The corresponding stocks of worms in the ordered matrix were thawed and genomic DNA isolated by standard protocols for PCR amplification (i.e. 5H 1-12, 10B 1-12; see Figure 6.6). Figure 6.6B shows the PCR screening of DNA derived from 5H 1-12 and 10B 1-12 with the *pbm-1* specific primers ZK1-4 and the robust amplification of a ~3.0kb wild type fragment in the majority of cases. In lane 5H4 the ~1.0kb deletion fragment is present in quadruplicate. Figure 6.6C shows the PCR amplification from

## Figure 6.6 : Identification of an EMS deletion within a mutant library

The term “positive address” refers to the reproducible amplification of a deletion fragment smaller than the wild type from genomic DNAs derived from a frozen stock of polyclonal nematodes. The “dimension” of this address refers to the location within a frozen ordered matrix.

### A. Identification of a first dimension positive address

In most lanes the amplification of a wild type fragment of 2.9kb is seen. The failure of the ZK1-4 primers to amplify a wild type fragment in all cases is likely due to the quality of the genomic DNA template that diminishes as the number of freeze-thaw cycles increases (Karen Thijssen, personal communication). The screen is carried out in quadruplicate to distinguish deletion fragments from spurious PCR products. In lanes 5H (3/4 positive deletion fragments) and 10B (2/4) the *pbm-1* primers amplified a deletion fragment of approximately 1.0kb (indicated by arrows and an asterisk). This suggested that the corresponding stock of frozen worms from which the genomic DNA was derived contained an animal(s) with a large (~2.0kb) 5' deletion in the *pbm-1* gene.

### B. Identification of a second dimension positive address

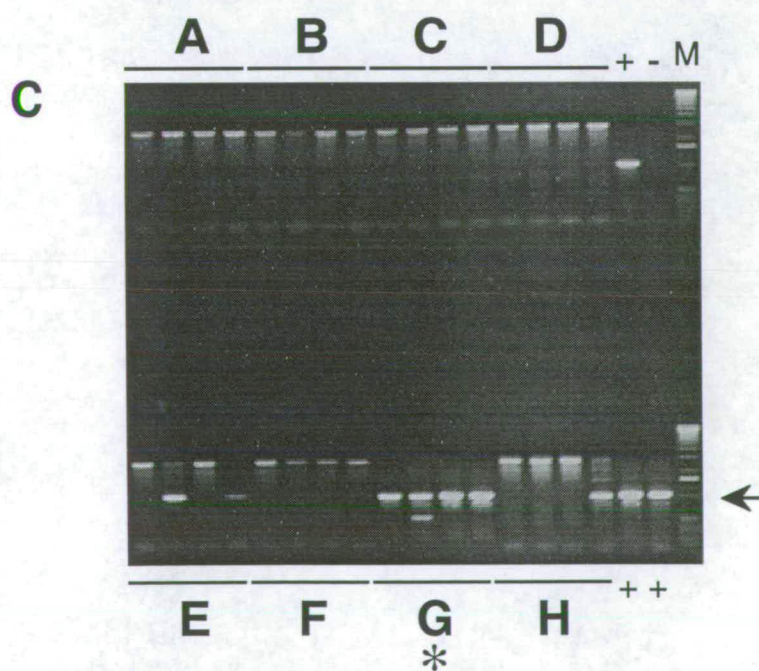
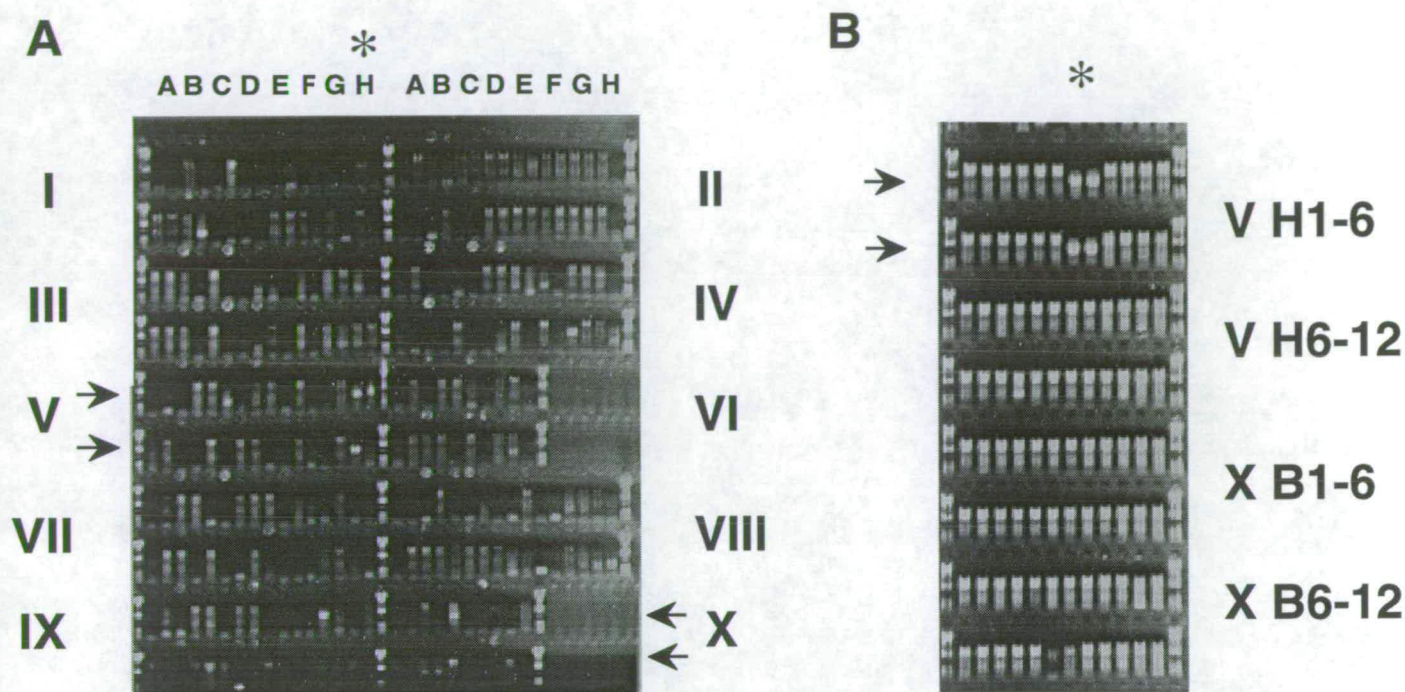
Genomic DNA was isolated by standard protocols for PCR amplification from the addresses 5H and 10B (i.e. 5H 1-12, 10B 1-12) and screened for deletions within the *pbm-1* gene. The majority of lanes amplify a wild type 3.0kb fragment from the *pbm-1*. In lane 5H4 the ~1.0kb deletion fragment present in quadruplicate is indicated.

### C. Identification of a third dimension positive address

Genomic DNA was isolated from the 5H4 addresses (i.e. 5H4 A-H) and 5H4G amplified in quadruplicate the ~1.0kb deletion band.

+ is a positive control using genomic DNA from the 5H4 address and  
- is a negative control with no DNA added.





genomic DNA derived from the 5H4 addresses (i.e. 5H4 A-H) and the detection of the 1.0kb deletion band in quadruplicate in 5H4G. The frozen stock of worms at 5H4G in the ordered deletion library was thawed and transferred to a 10cm seeded NGM agar plate and grown for 2 days at 25°C. These cultures are oligoclonal and thus it is unknown what percentage of the animals that survived the thawing procedure contain the ~2.0kb deletion within *pbm-1* (Plasterk, 1995). To ensure the deletion mutation in *pbm-1* was maintained, hermaphrodite worms were picked to 500 X 2cm seeded NGM plates and cultures of increasing complexity were established, i.e. 10 plates of 10 worms, 5 of 20 worms, 5 of 50 worms and a plate with the remaining population of animals. The parental hermaphrodites from the 500 individually seeded NGM plates, that importantly had progeny and would thus segregate the deletion, were transferred to worm lysis buffer and genomic DNA prepared for PCR amplification (as described in section 2.14.4). Figure 6.7B shows that a single founder hermaphrodite (line 58), contained the 5' deletion within the *pbm-1* gene whereas the remaining lines amplified the wild type fragment. Following identification of the *pbm-1* deletion mutant in this series, the small number of F<sub>1</sub> progeny of the mutant line (11) were transferred to individual NGM plates. This reduced fertility of EMS mutagenised nematodes has been observed previously (Jansen et al., 1997). These worms were PCR genotyped when they had laid viable progeny to ensure the mutation had segregated to the small number number of animals. Figure 6.7C shows that from the 11 lines, three amplified a ~3.0kb fragment and were thus homozygous wild type for *pbm-1* whereas the remaining

**Figure 6.7 : Identification of an EMS deletion within the *pbm-1* gene**

A. The 5.1kb *pbm-1* genomic locus and location of the ZK1-4 primers that amplify a wild type 2961bp fragment from the 5' end of the gene.

The frozen stock of worms at position 5H4G in the ordered deletion library identified by PCR screening were thawed and transferred to a 10cm seeded NGM agar plate and grown for 2 days. Hermaphrodite worms were then picked to 500 seeded NGM plates. When viable F1 progeny were visible on these plates, genomic DNA was prepared from the 500 parental hermaphrodites.

B. A single founder hermaphrodite (line 58) marked by an asterisk, contained the 5' deletion of *pbm-1* whereas the remaining lines amplified the wild type ~3.0kb fragment. + is a positive control using genomic DNA from the 5H4G address.

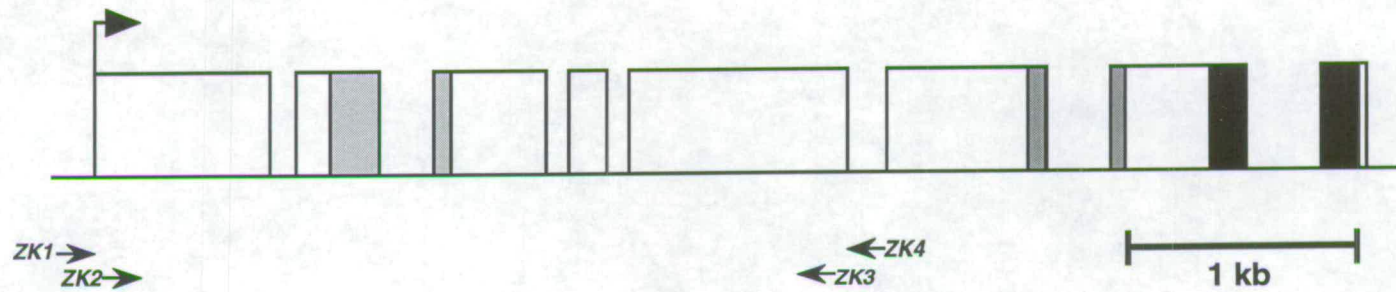
C. Following identification of the *pbm-1* deletion mutant in this series, the small number of F<sub>1</sub> progeny (11) of the mutant line 58 were transferred to individual NGM plates. These worms were PCR genotyped when they had laid viable progeny to ensure the mutation had segregated to the small number of animals. From these 11 lines, three amplified the 3.0kb fragment and were thus homozygous wild type for *pbm-1* whereas the remaining eight lines segregated the mutation. As the ~1.0kb deletion band has a selective advantage in the PCR amplification it was not possible to say whether these lines were heterozygous or homozygous for the mutation.



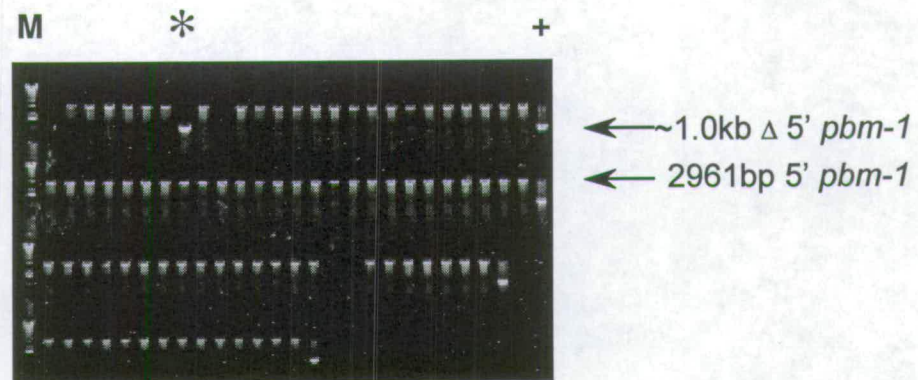
eight lines contained the mutation. As the deletion band has a selective advantage in the PCR amplification it was not possible to say whether these lines were heterozygous or homozygous for the mutation. Subsequently, internal primers to the deletion allowed discrimination of wild type and deletion mutant alleles to allow unambiguous genotyping of the worms and are described in section 6.2.3.

To characterise the nature of the deletion at the 5' of the *pbm-1* gene, the ~1.0kb *pbm-1* deletion fragment was TA-cloned and sequenced from both ends. Alignment of the sequenced 1088bp deletion fragment with the wild type genomic locus (from cosmid ZK783, Accession No. U13646) revealed an 1873bp deletion at the 5' end of the *pbm-1* gene. Figure 6.8A shows the eight exon gene and the location of the ZK1-4 5' primers used to isolate the deletion mutant and the 1.8kb deletion in *pbm-1*. The deletion joins the 5' end of exon 1 to the extreme 3' end of exon 4, just upstream of the splice donor of this exon (see Figure 6.8A). Importantly, this deletion completely removed exons 2 and 3 that encode the MBD of PBM-1. The 1.8kb deletion generates a +1 frameshift mutation in the *pbm-1* gene such that the resulting out of frame mRNA encodes a severely truncated protein not containing any of the conserved motifs in PBM-1 and is therefore considered a null allele.

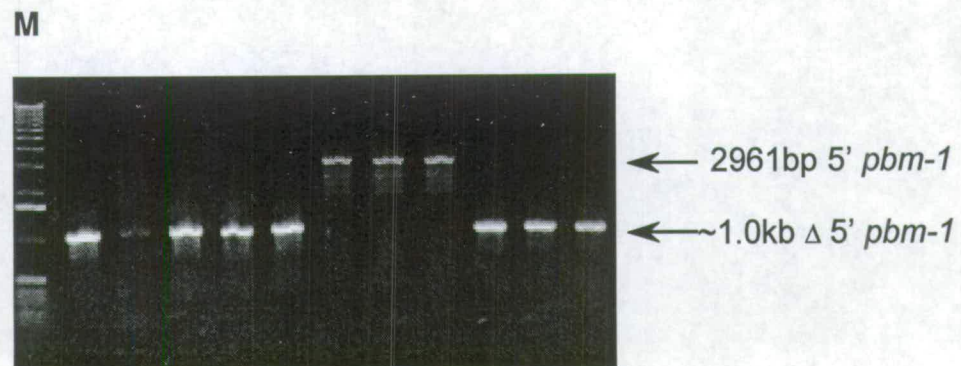
**A**



**B**



**C**



**Figure 6.8 : Establishment of a strain homozygous for the EMS deletion within the *pbm-1* gene**

**A. Characterisation of the EMS deletion at the 5' end of the *pbm-1* gene**

The *pbm-1* genomic locus and location of the ZK1-4 primers that amplify a wild type 2961bp fragment from the 5' end of the gene and a 1088bp deletion allele. PCR with the ZK1 and ZK6 primers then amplification with the nested primers ZK2 and ZK5 generates a 556bp fragment from the wild type locus but no fragment from the deletion allele.

The deletion fragment was sequenced and found to delete an 1873bp fragment from the 5' end of the *pbm-1* gene. This joins the 5' end of exon 1 to the extreme 3' end of exon 4, just upstream of the 5' splice donor of this exon shown in bold (see Table 3.1 for exon4-intron4 junction in *pbm-1*). This deletion completely removed exons 2 and 3 that encode the MBD of PBM-1. The 1.8kb deletion generates a +1 frameshift mutation such that the resulting mRNA encodes a severely truncated protein not containing any of the conserved motifs in PBM-1 and is therefore considered a null allele.

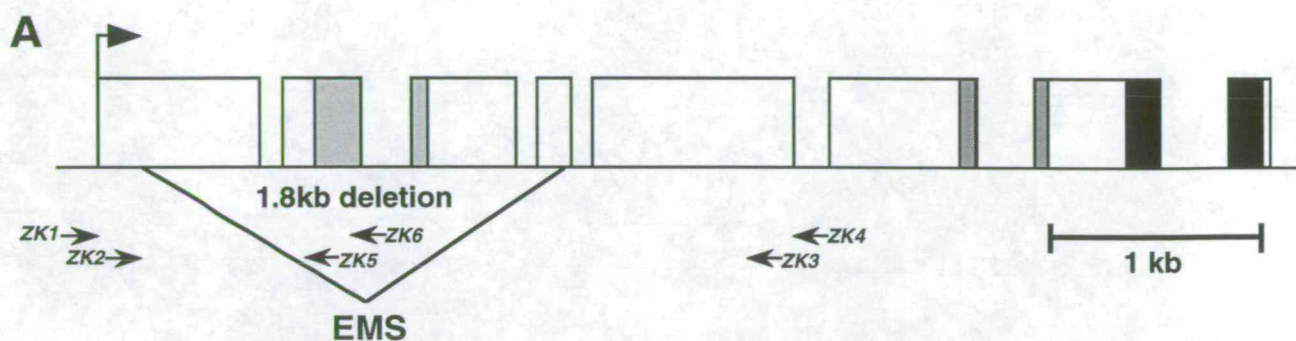
**B. A homozygous *pbm-1* strain for the deletion could not be identified**

The *pbm-1* deletion was isolated as a heterozygote from the deletion mutant library and a homozygous line could not be identified. The failure to isolate a worm strain homozygous for the *pbm-1* deletion suggested the mutation might be a recessive lethal or alternatively was linked to a second site mutation that was recessive lethal. Shown in the figure are 12 worm lines either homozygous wild type (+/+) or heterozygous for the *pbm-1* deletion (+/-). The PCR genotyping scheme used primers ZK1-4 to determine if the strain was wild type (2961bp wt) or harboured the deletion (1088bp Δ). PCR amplification with primers ZK1/6 and then nested PCR with primers ZK2/5 that are internal to the 1.8kb EMS deletion determined whether the strain was heterozygous (presence of a 556bp product) or homozygous (absence of a 556bp product) for the deletion allele.

**C. Isolation of a strain homozygous for the *pbm-1* deletion after outcrossing**

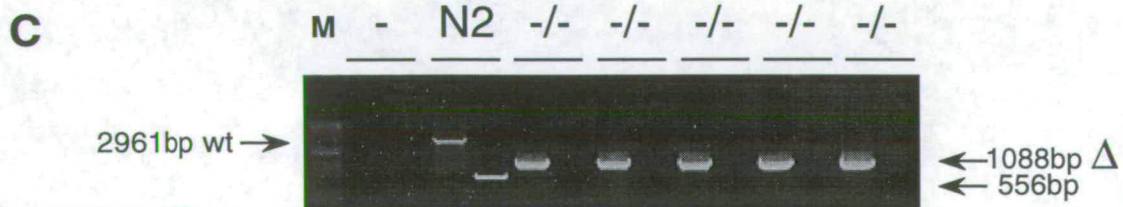
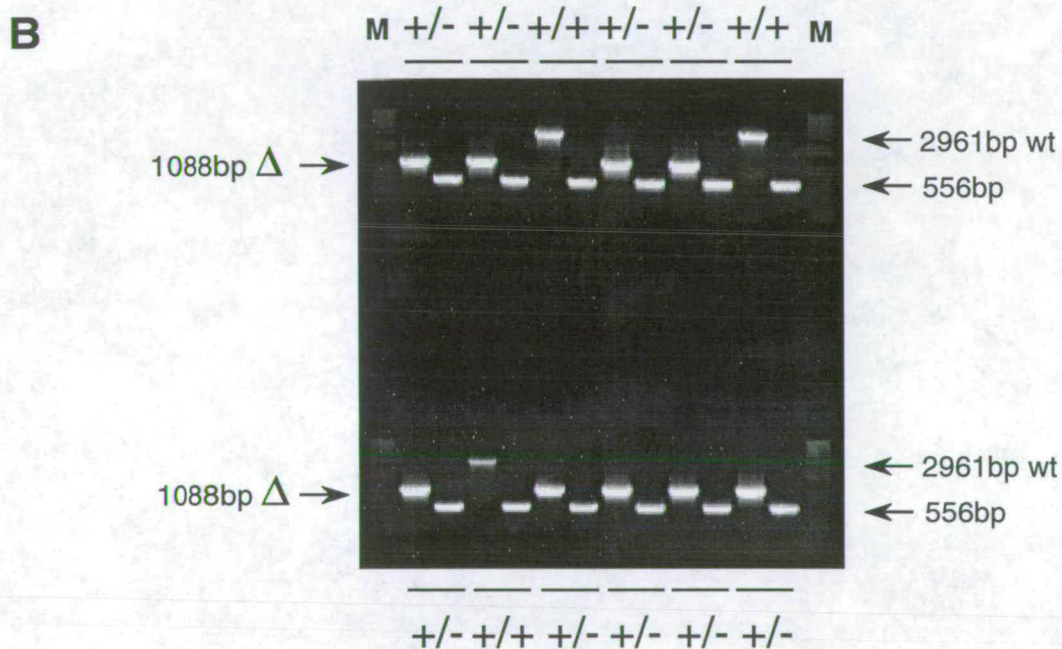
To remove second site mutations to stabilise the strain and facilitate phenotypic analysis, extensive outcrossing to an unmutagenised genetic background was carried out using chromosome specific markers (see section 6.2.3). At outcross three a strain homozygous for the ~1.8kb deletion was identified. Outcrossing to the unmutagenised background was carried out five times in total and the strain was deposited at the CGC, University of Minnesota, USA as *pbm-1* (*qa 1700*). The hermaphrodite gross phenotype in the *pbm-1* (*qa 1700*) strain was wildtype.

- is a negative control without added DNA. + is a positive control amplifying a wild type 2961bp and 556bp *pbm-1* fragments from genomic DNA isolated from the standard Bristol N2 strain. Adjacent to this are five lines homozygous for the deletion in *pbm-1*.



Exon 1                      Exon 4    Intron 4

TCAACTTCTG [..... 1873 bp deletion .....] AAAAG/gtaatt...



### 6.2.3 Outcrossing of the *C. elegans* deletion mutant

Internal primers to the 1.8kb deletion of *pbm-1* were designed to facilitate PCR genotyping to distinguish wild type and deletion alleles (see Figure 6.8A).

#### *pbm-1* specific primers

ZK5 : 5'-AATTGAAGCTGATGATGTGG-3'

ZK6 : 5'-ATTGTTTCATCATATTGCTCG-3'

The *pbm-1* deletion was isolated as a heterozygote from the mutant library and a homozygous line could not be identified (Figure 6.8B). This suggested that there might be selective pressure against isolation of a homozygous *pbm-1* strain. However, the EMS in vivo mutagenesis generates numerous, extraneous second site mutations that need to be removed before phenotypic analysis can be carried out. One straightforward explanation for the failure to isolate a *pbm-1* homozygote could be the result of linkage to a gene heterozygous for an EMS-deletion that is lethal in homozygous gene dosage.

To remove second site EMS mutations, extensive outcrossing to an unmutagenised genetic background was carried out using chromosome-specific markers. The *pbm-1* gene is located on *C. elegans* chromosome III so the chromosome III-specific markers *dpy-17* and *unc-32* were chosen to ensure outcrossing had occurred and to improve the mating efficiency. *Dpy-17;unc-32* (DpyUnc) homozygotes are short, severely coiled nematodes that move little as adult hermaphrodites and thus provide an easily distinguishable, recessive visible marker. The outcrossing scheme is shown in Figure

6.9. This procedure was repeated five times and a homozygous *pbm-1* strain was established after three outcrosses (see Figure 6.8C) suggesting the removal of a linked recessive lethal mutation. The five times outcrossed strain was deposited with the Caenorhabditis Genetics Center (CGC, University of Minnesota, USA) as *pbm-1 (qa1700)*.

Total RNA was isolated from the outcrossed *pbm-1 (qa1700)* strain, cDNA synthesised and RT-PCR analysis carried out to characterise the nature of the mature mRNA produced from the deletion mutant locus. This was accomplished by using primers specific to exons 1 and exons 5, that span the 1873bp deletion in the *pbm-1* mutant and amplified a predicted 2346bp fragment from wild type (Bristol N2) cDNA.

CEZKEXON1/F : 5'-ATGAGTGATAACTCATCTAATC-3'  
CEZKEXON5/R : 5'-AGCTTCTTGATGAAGATCTCCG-3'

In contrast, using cDNA derived from total RNA isolated from the *pbm-1 (qa1700)* strain, a 748bp fragment was amplified. Sequencing of these fragment confirmed the prediction that the *pbm-1 (qa1700)* allele encoded an out of frame mRNA that is prematurely truncated removing the PHD-zinc finger and bromodomain motifs. The allele is thus considered a null. The establishment of a strain homozygous for a null allele demonstrates that the *pbm-1* gene is therefore non-essential on the basis of not

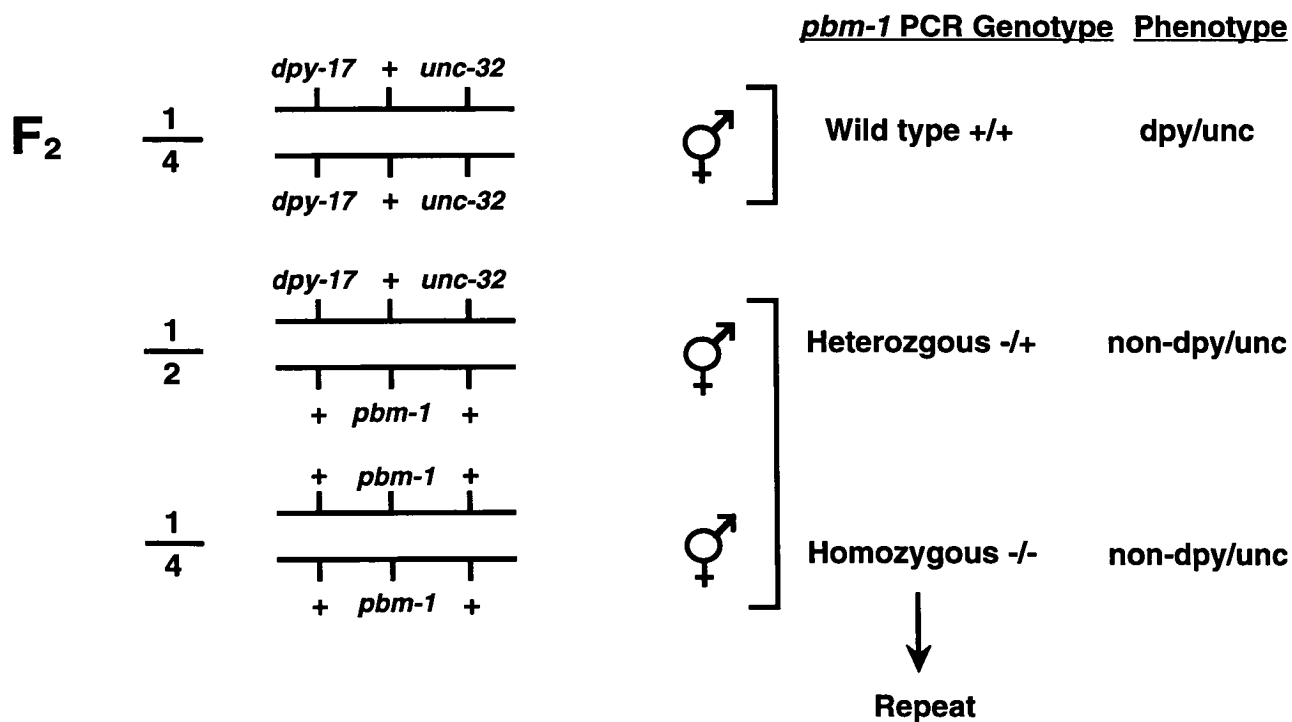
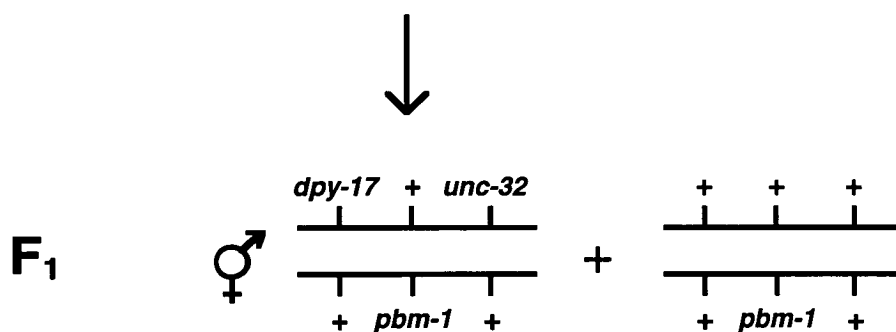
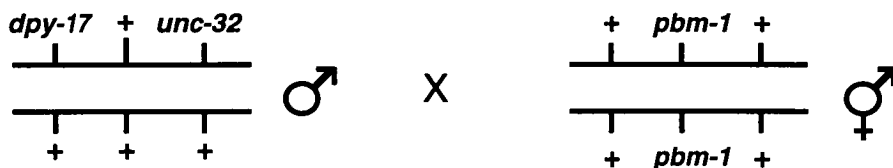
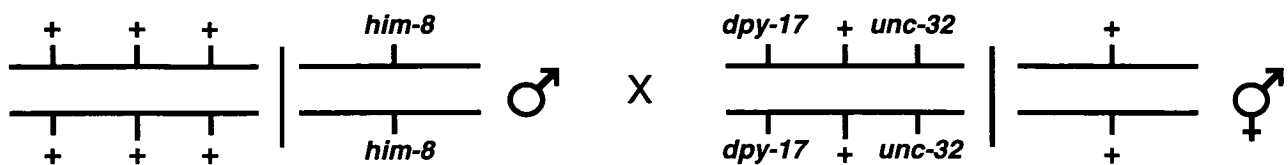
**Figure 6.9 : Outcrossing scheme to remove second site mutations from the *pbm-1* deletion strain**

The cis-linked (chromosome III)-specific markers *dpy-17;unc-32* were used in this outcrossing scheme to distinguish *self* and *cross*-progeny in the matings to ensure outcrossing to the unmutagenized background had occurred.

Adult hermaphrodites homozygous for *dpy-17* (*e164*) and *unc-32* (*e169*) III (SP471) were mated with an excess of young adult males *him-8* (*e1489*) IV. Male *cross*-progeny, i.e. from plates that contained phenotypically wild type worms as well as DpyUnc worms, were mated from the third outcross stage with adult hermaphrodites homozygous for the *pbm-1* deletion. Twenty L4 hermaphrodite progeny from this cross were seeded individually and allowed to self-fertilise. F<sub>2</sub> progeny from lines heterozygous for the *pbm-1* deletion by PCR genotyping were allowed to self, PCR genotyped and the lines that were positive for the deletion were identified. This scheme was repeated five times and a homozygous *pbm-1* strain was established after three outcrosses suggesting the removal of a linked mutation that was recessive lethal. The five times outcrossed strain was deposited with the Caenorhabditis Genetics Center (CGC, University of Minnesota, USA) as *pbm-1* (*qa1700*).

Outcrosses one and two were performed by mating hermaphrodites heterozygous for the *pbm-1* deletion with *dpy-17;unc-32* heterozygous males and as outlined above following the segregation of the deletion allele by PCR genotyping of their progeny.





being mutable to a recessive lethal. Furthermore, the hermaphrodite gross phenotype in the homozygous *pbm-1* mutant was indistinguishable from wild type.

### 6.3 Conclusions

This chapter describes a reverse genetic approach for target-selected gene inactivation of the *pbm-1* gene on *C. elegans* chromosome III. A 1.8 kb EMS induced deletion at the 5' end of the *pbm-1* gene was isolated from a mutant library. Using the chromosome III-specific markers *dpy-17* (*e164*) and *unc-32* (*e169*), the *pbm-1* deletion strain was outcrossed extensively into an unmutagenised genetic background. A strain homozygous for the deletion, *pbm-1* (*qa1700*) was established and RT-PCR analysis from this homozygous mutant strain confirmed that the deletion was a null allele. The establishment of a *C. elegans* strain that is homozygous for a severe deletion allele demonstrates that the *pbm-1* gene is non-essential for viability.

**CHAPTER SEVEN: DISCUSSION**

## 7.1 Summary of Results

1) A *C. elegans* gene was identified that exhibited homology to the methyl-CpG binding domain of the MBD protein family. Intriguingly, a cDNA derived from this locus encoded a protein containing PHD-zinc finger and bromodomain motifs found in chromatin-associated transcriptional regulators. The *C. elegans* gene was termed PHD-zinc finger, bromodomain and methyl-CpG binding domain gene-1 (*pbm-1*).

2) A mammalian ortholog of the nematode *pbm-1* gene was isolated. mPBM1 was found to be alternatively spliced with three major RNA isoforms detectable in a range of somatic and tissue cultured cell lines.

3) During the identification of mPBM1 a partial cDNA, mPBM2 was identified that potentially encoded a protein highly homologous to the C-terminus of the *C. elegans* PBM-1 protein. Corresponding paralogous cDNAs (hPBM1 and hPBM2) were identified in human tissues suggesting the existence of a gene family in mammals.

4) The MBD regions from *C. elegans* PBM-1 and mPBM1 were produced as recombinant proteins. No affinity of these domains for methylated DNA was detected in a series of in vitro binding studies. Furthermore, no binding of these MBD domains to hemi-methylated DNAs, CpG/TpG, m<sup>5</sup>CpG/TpG mismatches and a probe containing multiple TpG sites was observed.

5) The full length mPBM1 protein fused to the GFP reporter was found to be nuclear localised in a diffuse, granular pattern in mouse cells. The mPBM1 GFP signal did not overlap foci derived from the methylated mouse major satellite, strongly suggesting that mPBM1 does not associate with methylated domains in vivo. Furthermore, the MBD domain is not responsible for the distinct sub-nuclear localisation pattern of mPBM1. Deletion of the PHD-zinc finger/bromodomain however abolishes the uniformly distributed pattern of fluorescent foci throughout the nucleoplasm implicating these regions in the wild type localisation of the mPBM1 protein in mouse nuclei.

6) GAL4-mPBM1 tethered to the promoter of a reporter gene was found to exert a modest dose-dependent transcriptional repression in mouse L929 fibroblasts. Expression of a GAL4-mPBM1 fusion protein however was found to result in the activation of a reporter gene lacking GAL4 binding sites, suggesting that the silencing on the reporter containing upstream GAL4 binding elements was a specific effect arising from the recruitment of the mPBM1 protein.

7) The hPBM1 gene has been mapped by a number of independent approaches to 2q23. Disruption of genes encoding PHD-zinc finger proteins have been implicated in leukaemogenesis. The mapping of hPBM1 to 2q23 makes it a candidate for disruption in the cases of leukaemia involving rearrangements and deletions on the distal arm of human chromosome 2.

8) An EMS mutant library was screened to identify deletion events within the *pbm-1* gene. The deletion allele was shown to be a null by RT-PCR analysis. A *C. elegans* strain *pbm-1* (*qa1700*) was isolated that was homozygous mutant for this loss-of-function deletion. The *pbm-1* gene was not essential for viability.

## 7.2 The PHD-zinc finger, Bromodomain and MBD (PBM) protein family

The *C. elegans* *pbm-1* gene was identified in a search of the sequence databases for proteins with homology to the methyl-CpG binding domain. An alignment of the MBD of MBD1 with the MBD of PBM-1 over 40 amino acids showed 30% identity and an overall similarity of 52% (see Figure 3.13). Inspection of the extreme C-terminus of PBM-1 revealed the presence of a PHD-zinc finger and an immediately adjacent bromodomain strongly suggesting a relationship with gene regulation (Figures 3.11 and 3.12). The identification of an MBD-containing protein in an organism apparently devoid of genomic methylation is not without precedent as an EST has been identified in *Drosophila melanogaster* that is highly homologous to the mammalian MBD2/3 proteins (Susan Tweedie, personal communication). The MBD of *Drosophila* MBD2/3 contains a large insertion and does not form a complex with a methylated probe in vitro. Furthermore, a cDNA has been identified in *Drosophila* that shares sequence identity to the fission yeast methyltransferase homolog *pmt1* (Wilkinson et al., 1995) and the mammalian cytosine-5 methyltransferase family member, Dnmt2 (Okano et al., 1998a) (Susan Tweedie, personal communication). These findings demonstrate that *Drosophila* encodes proteins homologous to the mammalian DNA methylation machinery and yet like *C. elegans* no m<sup>5</sup>C has been

detected in the fly genome (Rae and Steele, 1979). In marked contrast, no additional MBD-containing genes or sequences with homology to the DNA (cytosine-5) methyltransferase family members were detected in the *C. elegans* genome.

A full length mouse cDNA termed mPBM1 was identified that was highly homologous to the *C. elegans* PBM-1 protein and importantly had the same modular organisation of motifs, i.e. the N-terminal MBD and C-terminal PHD-zinc finger/bromodomain regions. During the identification of mPBM1, a partial cDNA (mPBM2) was identified that potentially encoded a protein highly homologous to the C-terminus of the *C. elegans* PBM-1 protein. Corresponding partial cDNAs (hPBM1 and hPBM2) were identified in human tissues suggesting the existence of a family of PHD-zinc finger, Bromodomain and MBD (PBM)-containing proteins in mammals.

Two separate subclasses of gene homology : orthology and paralogy have been described by Fitch (1970). Orthologous genes (ortho = exact) have been defined as homologous genes in *different* species that encode proteins with the same biochemical function that have diverged from their common ancestral gene as a result of speciation and separate evolution. Paralogous genes (para = in parallel) are homologous genes in the *same* species that arise through gene duplication encoding proteins with related but non-identical biochemical functions. In this scheme, the mPBM1 and mPBM2 genes are paralogous to one another and paralogous genes comprise the many gene



families that have been identified. However, as pointed out by Chervitz et al., (1998) in comparing the entire protein complements of the completely sequenced *C. elegans* and *Saccharomyces cerevisiae* genomes, orthology is not a one-to-one relationship as a unique gene in one species (in this case the *C. elegans pbm-1* gene) may be the ortholog of a gene family in another species (here the mouse PBM family). Chervitz et al., (1998) arbitrarily chose to cluster worm and yeast genes in groups that had BLASTp values of  $p < 10^{-50}$ . On the basis of this criteria, the completely sequenced worm PBM-1 and mPBM1 cDNAs were considered orthologous pairs in this study (BLASTp value of  $p < 10^{-79}$ ). However, the assignment of mPBM1 as the ortholog of the *C. elegans* PBM-1 protein cannot be definitive in the absence of complete genome sequence information in the mouse, as a gene encoding a protein with a higher sequence identity to the ancestral *pbm-1* gene could reside in the unsequenced portion of the mouse genome.

### **7.3 The MBD regions of the PBM proteins do not bind methylated DNA**

Recombinant proteins comprising the MBD regions from the worm PBM-1 and mouse PBM1 proteins were expressed in *E. coli*. Purified proteins were found to express well with minimal proteolytic degradation. In order to test whether these MBD regions were capable of binding to a methylated template, a series of bandshift and southwestern assays were performed. A positive control (MeCP2, MBD1) could bind readily to a methylated template whereas no binding of the MBD domains from the worm PBM-1 and mouse PBM1 proteins could be detected. In further

experiments using probes containing multiple TpGs, hemi-methylated DNAs, CpG/TpG and m<sup>5</sup>CpG/TpG mismatch probes no binding was observed whereas positive controls were found to form specific complexes with these DNAs. The sequence similarity between the MBD region from the worm PBM-1 and mouse PBM1 proteins is reduced when compared with the MBD family members known to complex with methylated probes (MeCP2, MBD1, MBD2 and MBD4; see Figure 3.13).

An NMR structure for the MBD of human MeCP2 in solution has been deduced and a number of critical residues within a conserved hydrophobic pocket are proposed to mediate the contacts between the MBD and m<sup>5</sup>C in DNA (Wakefield et al., 1999). The MBD of MeCP2 is proposed to form a wedge-shaped structure with four N-terminal anti-parallel  $\beta$ -sheets (A-D) forming one side of this wedge and a C-terminal  $\alpha$ -helix forming the other. The authors suggest that Tyr123 is juxtaposed on the surface of the protein with the charged residues Arg111, Asp121 and Arg133 nearby critical for making contact with the m<sup>5</sup>CpG dinucleotide (see Figure 3.13). The Asp121 residue is known to be absolutely required for binding of the MBD of MeCP2 to methylated DNA as site-directed mutagenesis to a glutamate residue abolishes binding (Wakefield et al., 1999; Andrew Free, personal communication). Of these residues thought to be critical for binding to methylated DNA within the MBD of MeCP2, Arg111 is present in the PBM-1 and mouse PBM1 MBDs while Asp121 is present in the PBM-1 protein however a glutamate residue exists in this position in

mPBM1. The conserved Arg133 is conserved in the mPBM1 MBD however is “mutated” to serine in the worm PBM-1 sequence. Strikingly, the Tyr123 residue is not present in either the worm PBM-1 or mPBM1 MBDs. Interestingly the adjacent residues at positions 124 and 125 contain conserved aromatic residues (see Figure 3.13). These findings considered with the inability to detect a complex between the recombinant MBDs of PBM-1 and the mouse PBM1 proteins and methylated DNA strongly suggests that the methyl-CpG binding domain in these proteins is non-functional. However, the experiments carried out in this study do not exclude the possibility that the MBD regions mediate sequence specific DNA-binding in these proteins to other target sequences. Alternatively, on the basis of sequence conservation between the nematode PBM-1 and mammalian mPBM1 MBDs these motifs may serve as RNA-binding motifs or potentially as protein-protein interaction motifs within these proteins.

The mPBM1 cDNA was the product of a complex series of alternative splicing events located adjacent to the MBD (see Figure 3.7). RT-PCR detected a 95 amino acid splicing event N-terminal of the MBD and spliced and unspliced RNAs were detected in somatic and tissue cultured cells. Interestingly, alternative splices adjacent to DNA-targeting domains have been found to modulate binding in different isoforms and the basic helix-loop-helix (bHLH) family of transcription factors provide good examples of this effect. The bHLH-ZIP transcription factor microphthalmia (Mi) binds DNA as a homodimer (Hemesath et al., 1994). Mi has two isoforms that differ

by an alternatively spliced insert of 6 amino acids N-terminal of the basic DNA binding region and the presence of the insert was found to enhance DNA binding (Hemesath et al., 1994). By analogy with the bHLH family of transcription factors, the PBM-1 and mPBM1 proteins may bind to DNA as homodimers and the expression of the MBD regions alone would not allow dimerisation to occur to facilitate binding. One possibility that should be tested is that the different splice variants encode proteins that either stimulate or inhibit DNA binding via the MBD region in mPBM1. A number of techniques exist to determine the consensus target sequence of a suspected DNA binding domain. A binding site selection approach with an oligonucleotide containing a randomised internal region (Oliphant and Struhl, 1987) or whole genome PCR (Kinzler and Vogelstein, 1989) could be used to determine whether *C. elegans* PBM-1 and mouse PBM1 proteins have affinity for DNA. Bulyk et al., (1999) have reported a novel double stranded DNA array for investigating DNA-protein interactions and have shown that the immobilised DNA can be covalently modified by methylation allowing recognition sequences containing post-synthetic modifications to DNA to be screened for interactions.

#### **7.4 Localisation of the mPBM1 protein**

The mPBM1 protein has been localised in mouse fibroblasts by fusing the full length protein to a green fluorescent protein marker. The protein gave a distinct granular localisation pattern throughout the nucleoplasm but appeared to be absolutely excluded from the nucleoli. The mouse major satellite has been reported to contain

some 50% of all m<sup>5</sup>C in the mouse genome. The mouse satellite is present in pericentromeric heterochromatic regions which stain strongly with the DNA dye DAPI (Miller et al., 1974). Importantly, no co-localisation of the mPBM1-GFP signal with the DAPI signal was observed. This finding in vivo that the mPBM1 does not target to heavily methylated foci is in agreement with the in vitro binding studies which failed to detect an affinity for a methylated template. Furthermore, deletion of the N-terminal MBD region did not perturb localisation of the mPBM1 protein in the nucleus however the removal of the entire C-terminus containing the PHD-zinc finger and bromodomain regions resulted in a dramatically altered localisation pattern with foci of intensely labelled signal as opposed to the diffuse pattern of the wild type protein.

Intriguingly, similar results were observed by Rinderle et al., (1999) studying localisation of another PHD-zinc finger protein, the putative autoimmune regulator, *AIRE*. APECED (autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy; OMIM 240300) is an autosomal recessive human genetic disorder and arises through production of autoantibodies leading to a potent autoimmune response directed at numerous organs and autoantigens (Campbell and Milner, 1993). Nagamine et al., (1997) reported the positional cloning of *AIRE*, a putative transcription factor by virtue of possessing two PHD-zinc fingers, homologous to the single C-terminal mPBM1 zinc finger (see Figure 3.9). Mutational analysis in Finnish and Swiss APECED families identified a predominant, nonsense mutation

(R257X) in *AIRE* which results in a Arg→STOP nonsense mutation leading to truncation of the C-terminal PHD-zinc fingers. This suggested that the molecular aetiology of the APECED disorder might arise through mislocalisation of AIRE due to the absence of the zinc-finger domains. The nuclear sub-localisation of AIRE was strikingly similar to that of mPBM1, with immunofluorescence revealing the protein to be distributed in discrete, punctate speckles throughout the nucleoplasm (Rinderle et al., 1999), reminiscent of the localisation of full length mPBM1 fusions (Figure 4.1A). However, mutational analysis of AIRE deleting the two PHD-zinc fingers (corresponding to the major Finnish R257X mutation) lead to a dramatically altered nuclear localisation where the mutant protein concentrated in discrete foci as opposed to its normally diffuse pattern. This mimicked the mPBM1 mutant protein that lacked the C-terminal PHD-zinc finger/bromodomain that was also found to be present in intensely labelled foci following deletion of a domain containing this highly conserved zinc finger region. The perturbed nuclear localisation of mPBM1 $\Delta$ PHD/BrD and that of the AIRE R257X mutant protein may thus arise through a failure to associate with nuclear proteins that confer the wild type diffuse, granular pattern via PHD-zinc finger regions.

Importantly, the work by Rinderle et al., (1999) on AIRE localisation did not test directly by site-directed mutagenesis of the zinc finger whether this domain mediates targeting to specific nuclear sub-domains. The diffuse dot-like distribution of full length mPBM1 is reminiscent of the pattern observed with other nuclear proteins,

particularly the promyelocytic leukaemia (PML) protein (Dyck et al., 1994) and the acute lymphoblastic leukaemia (ALL1) protein (Yano et al., 1997). In order, to investigate the structures containing the mPBM1 polypeptide, double-immunofluorescence experiments using antibodies specific to mPBM1 and PML or ALL1 would determine whether these activities co-localise.

These studies on mPBM1 localisation were carried out using a GFP marker and overexpressed levels of protein. For *in vivo* immunohistochemical detection of the mPBM1 protein a specific polyclonal antibody is required to determine whether the diffuse staining pattern seen in fibroblasts ectopically expressing a GFP-tagged fusion protein is reflective of the endogenous localisation pattern. A single mouse fibroblast cell line (EFS2) was used in this study to localise the mPBM1 activity and it is necessary to determine whether the requirement for an intact C-terminus to confer a diffuse granular localisation pattern is reflective in a range of transfected cell types.

A growing number of human genetic diseases have been identified that arise through mutation of PHD-zinc finger genes thought to encode general transcriptional regulators. In addition to the PHD-zinc finger containing protein AIRE found mutated in the APECED autoimmune disorder described above, mutations in the PHD-like finger of hATRX, a SNF2 family member has been shown to cause the X-linked developmental disorder ATR-X syndrome (Gibbons et al., 1997). Mutation and deletion of the PHD-zinc finger/bromodomain-containing CREB-binding protein



(CBP) have been found to be responsible for another developmental defect, Rubinstein-Taybi syndrome (RTS) (Petrij et al., 1995). Williams syndrome (WS) is a complex developmental disorder with multiple developmental defects caused by deletion of chromosome 7q11.23 and a PHD-zinc finger/bromodomain gene WSTF has been mapped to this WS deletion interval (Peoples et al., 1998; Lu et al., 1998). The ATR-X, RTS and WS human disease syndromes are characterised by developmental abnormalities and are thought to arise through a generalised deregulation of normal gene expression.

### **7.5 Mapping of the hPBM1 gene to 2q23**

Chromosomal translocations are frequently identified in patients with various forms of leukaemia (see Chapter 5). Clinical cases with a single cytogenetic abnormality facilitates the cloning of the breakpoints of the translocations leading to gene(s) identification involved in the molecular aetiology of leukaemia. The human PBM1 gene has been mapped by a number of approaches to human chromosome 2q23. Interestingly, a number of deletions and translocations of human chromosome 2q23 are known in human leukaemias.

Columbano-Green et al., (1990) have reported a cytogenetic analysis of bone marrow cells from a patient with acute nonlymphocytic leukaemia (ANLL) that had as the sole detectable cytogenetic abnormality the rearrangement t(2;14)(q23;q32.3). The only other t(2;14) reported was in a case of chronic myeloid leukaemia (CML)

(Sessarego et al., 1979). The position of hPBM1 gene on chromosome 2 is of interest because a number of cases of ANLL have been reported following deletions at 2q23 (Arthur et al., 1987) and like the study by Columbano-Green et al., (1990) the patients responded poorly to therapeutic intervention suggesting deletions/rearrangements of 2q23 are indicative of a poor clinical prognosis.

Thus, translocations and deletions involving 2q23 have suggested the existence of oncogenes or genes involved in controlling haemopoiesis at this locus. Transcription and developmental regulation are key elements in tumourigenesis where transcriptional regulators that control many downstream targets are thought to be crucial to the development of cancer by virtue of aberrantly regulating numerous loci. The mapping of hPBM1 to 2q23 makes it a candidate for disruption in the cases of leukaemia involving rearrangements and deletions on the distal arm of human chromosome 2.

The report of a deletion specifically at 2q23 is the most convincing evidence that disruption of gene(s) at this chromosomal location are responsible for transformation and not simply the result of mutation of genes on partner chromosomes that give rise to the neoplasia. However, the cytogenetic abnormality specifically at 2q23 is a contiguous gene deletion disorder where many loci that reside within this chromosomal interval could be responsible for the leukaemogenesis. In order to determine whether the hPBM1 gene was involved in the cases of leukaemia at

chromosome 2q23, genomic DNA from patients who are cytogenetically t(2;14)(q23;q32.3) could be analysed by Southern blot analysis to see if rearrangements existed within the gene. If rearrangements were detected then the breakpoint should be cloned to 1) characterise the nature of the deletion (intronic or exonic) within the hPBM1 gene and 2) to identify partner genes on chromosome 14. The first step would be to make somatic cell hybrids to have a source of genomic DNA to generate t(2;14)-specific cosmid libraries. This could be carried out by fusing viable leukaemic cells from t(2;14) patients to HPRT<sup>-</sup> Chinese hamster cells under selection for resulting hybridoma events. These hybrids could then be screened for the derivative (der) fusion chromosome 14 segregating from the normal chromosome 14 and the other derivative chromosome, der(2) by using Southern hybridisation and STS mapping. Screening of a cosmid library made with genomic DNA from the der(14) somatic cell hybrid with 5' and 3' hPBM1 probes would isolate chimaeric clones spanning the breakpoint. This junction cosmid spanning the breakpoint could then be exon trapped and chromosome 14 cDNAs identified. Chromosome 14-specific sequences that give an rearranged Southern pattern could then be used in RT-PCR on RNA from t(2;14) patients to characterise the nature of the oncofusion mRNA in the t(2;14) patients.

## **7.6 Transcriptional regulatory properties of the mPBM1 protein**

The mPBM1 protein sequence contains several motifs present in established transcriptional regulators. In this study, a fragment of mPBM1 fused to the

heterologous DNA-binding domain of GAL4 was found to be capable of a modest dose-dependent repression when tethered to a reporter gene in vivo. Importantly, no reduction in the transcriptional competence of the reporter gene was observed without the specific binding of the mPBM1 fusion. However, the repressive effect of the mPBM1 fusion was weak, resulting in only a 50% reduction in activity of the reporter (see Figure 4.4). In order to better characterise the transcriptional regulatory properties of mPBM1, nested deletions are required and the assaying of the effects of these constructs on a reporter would allow the mapping of transcriptional regulatory domains within the protein. Furthermore, the silencing of the reporter gene may occur by mechanisms involving steric hindrance or via protein-protein interactions. To test the possibility that the binding of mPBM1 on the reporter gene is inhibiting transcription directly by promoter occlusion, the GAL4-binding elements could be removed to a more distal location and assayed for repression in a distance-independent manner. Repression at a distance would exclude the possibility that sterically hindering the promoter is mediating the repressive effect. GAL4 fusions of the *C. elegans* PBM-1 cDNA could be tested for the ability to modulate a reporter gene in mammalian cells as has been demonstrated for *C. elegans* proteins that target the highly conserved RNA pol II transcriptional machinery (Batchelder et al., 1999). Testing the mPBM1 protein in yeast and *Drosophila* SL2 cells for the ability to modulate transcription would further strengthen a mechanism operating through repression of the conserved transcriptional machinery (Pengue and Lania, 1996).

Mapping of a repression domain and two hybrid screens for interacting proteins with this region would better characterise mediators of the silencing.

The mPBM1 cDNA was found to be alternatively spliced and functional diversification of transcriptional regulators is known to occur as a result of splicing events. For example, the AML1 gene encodes three isoforms (AML1 a-c) which possess distinct effects on transcription, differentiation and proliferation of myeloid cells (Tanaka et al, 1995). The repressive effect on the reporter gene by mPBM1 was observed with the mPBM1a isoform and it would be interesting to see whether the alternatively spliced mPBM1b and mPBM1c proteins (diagrammed in Figure 3.7) mediate distinct effects on transcription.

An interaction between the PHD-zinc fingers of Mi2 $\beta$  and the histone deacetylase, HDAC1 has been reported (Zhang et al., 1998). Treating the cells transfected with the mPBM1 fusion with an inhibitor of histone deacetylases such as trichostatin A (Yoshida et al., 1995) would establish whether deacetylation plays a part in the mechanism of transcriptional repression by mPBM1.

### **7.7 Potential protein interaction targets of the PBM proteins**

MBD-containing proteins have been implicated in the epigenetic control of gene transcription. CpG methylation can induce histone deacetylation, chromatin remodelling and gene silencing by assembly of a transcriptional repressor complex

containing the histone deacetylases, HDAC1/2 bound to the co-repressor, mSin3A (Nan et al., 1998; Jones et al., 1998; Eden et al., 1998). This silencing complex is assembled by the binding of MeCP2 to methylated DNA through its methyl-CpG binding domain (MBD) and the recruitment of the mSin3A/HDAC machinery via a separable transcriptional repression domain. These findings establish a link between the epigenetic phenomena of DNA methylation and histone deacetylation.

Interestingly, together with the MBD domains found in *C. elegans* PBM-1 and mouse PBM1, these proteins possess additional motifs implicated in epigenetic mechanisms of gene regulation : the PHD-zinc finger and bromodomain motifs. The presence of the PHD-zinc finger in the *Drosophila* trithorax and polycomb-like proteins which exert antagonistic effects on chromatin structure and gene expression has prompted the suggestion that this motif may be involved in the interaction between chromatinic proteins (Aasland et al., 1995). Subsequently, it has been demonstrated that the PHD-zinc fingers of Mi2 $\beta$  are required for interaction with the histone deacetylase HDAC1 within a large multiprotein complex termed NuRD that possesses histone deacetylase and chromatin remodeling ability (Zhang et al., 1998). The bromodomain motif is also found in protein complexes involved in chromatin remodeling such as in the GCN5 and TAF<sub>II</sub>250 intrinsic histone acetyltransferases (Brownell et al., 1996; Mizzen et al., 1996). Barlev et al., (1998) have identified a functional interaction between the bromodomain of GCN5 and the Ku70 subunit of the DNA-dependent protein kinase (DNA-PK) both in vitro and in vivo.

Phosphorylation of GCN5 by the DNA-PK holoenzyme resulted in an attenuation of GCN5s acetyltransferase potential.

To identify protein targets of the mPBM1 PHD and bromodomain motifs a yeast two-hybrid screen could be used to look for interacting proteins with this region in vivo. Barlev et al., (1998) have reported that in addition to GCN5, the bromodomains of the CREB-binding protein (CBP) and TAF<sub>II</sub>250 also interacted directly with the Ku70 component of the DNA-PK complex in vitro and in vivo. This would make the bromodomain of mPBM1 a strong candidate for interaction also, which could be tested for by a GST pull-down assay by immobilising a GST-Ku70 fusion on GSH-Sepharose beads and incubating with in vitro translated S<sup>35</sup>-labelled mPBM1. An in vivo interaction between mPBM1 and Ku70 could be tested for by transient co-transfection of mouse L929 cells with Myc-tagged Ku70 and HA-tagged mPBM1. Proteins immunoprecipitated with an anti-HA antibody could be probed for the presence of Ku70 with an anti-Ku70 antibody to see whether Ku70 can associate with mPBM1 in vivo. Similarly, an interaction between the mPBM1 PHD-zinc finger and histone deacetylases, as has been demonstrated for HDAC1 and the PHD-zinc finger of Mi2 $\beta$  reported by Zhang et al., (1998), could be tested for by a GST pull down assay with GST-HDAC1 and in vitro translated mPBM1 and a co-immunoprecipitation experiment of the type described above.



## 7.8 The *C. elegans* *pbm-1* gene is not essential for viability

In order to determine the function of a gene, an integration of information is required from a number of sources with respect to ortholog characterisation and the determination of a mutant phenotype. To obtain a *C. elegans* strain containing a loss-of-function mutant in the *pbm-1* gene, a deletion mutant bank was screened. An EMS-induced deletion allele was identified within the *pbm-1* gene and a *C. elegans* strain *pbm-1* (*qa1700*) was isolated that was homozygous mutant for this deletion. Sequencing of transcripts from this deletion locus confirmed that the allele was a null. The homozygous *pbm-1* mutant was viable and the *pbm-1* gene is therefore non-essential on the straightforward basis of not being mutable to a recessive lethal. No related genes were present in the *C. elegans* genome to account for functional redundancy within a multigene family where a related gene product might compensate for the loss-of-function *pbm-1* mutation.

Histone acetylation and deacetylation are known to play key roles in major differentiation events in *C. elegans* embryogenesis. Shi and Mello (1998) have shown that inhibition of the *cbp-1* gene which encodes a protein with significant sequence similarity to the mammalian CBP/p300 histone acetyltransferases resulted in differentiation defects in endoderm and mesoderm development. Interfering with the HDAC and RbAp48-related genes *hda-1*, *rba-1* and *rba-2* partially suppressed the differentiation defects seen in the *cbp-1* mutants. HDAC1 and RbAp48 are subunits of a mammalian histone deacetylase complex (Taunton et al., 1996) and may serve to

antagonise the activation function of CBP/p300. The PBM-1 protein shares homology with factors such as trithorax and polycomb-like involved in the developmental silencing of the clustered homeotic genes in *Drosophila*. Similarities exist between the silencing of the homeotic genes in *Drosophila* and the phenomenon of position-effect variegation where juxtaposition of a gene relative to heterochromatin results in transcriptional repression. Despite the large number of chromosomal rearrangements generated in *C. elegans*, there are few reports of position-effects and it has been suggested that chromatin organisation in *C. elegans* is more uniform than the mouse and *Drosophila* (Hodgkin, 1994).

Second-site modification of mutations is a powerful approach to identify proteins that either target a gene directly or interact with the protein product. In the absence of a discernible phenotype, the pathways in which the *pbm-1* gene operates cannot be isolated by traditional approaches such as the identification of enhancer or suppressor mutants. Gonadal microinjection of translational fusions between the *pbm-1* gene and GFP under the control of the endogenous promoter however would yield valuable information on the sites of expression which would aid in phenotype identification. Overexpression of PBM-1 in wild type worms could also generate a phenotype from which function might be ascertained. Antisera specific to PBM-1 would determine at what stage of development the *pbm-1* gene is expressed and in which cell types and to confirm that no protein expression is detectable in the *pbm-1* (*qa1700*) strain. As the lineage of every cell in the worm is known, the information

generated by localising the site of protein expression can be particularly informative and could be used to screen for mutations in other genes that affect the localisation of the wild type PBM-1 protein.

## REFERENCES

Aasland R, Gibson TJ, Stewart AF. (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci* 20:56-9

Ahringer J. (1997) Turn to the worm! *Curr Opin Genet Dev* 7:410-5

Alland L, Muhle R, Hou H Jr, Potes J, Chin L, Schreiber-Agus N, DePinho RA. (1997) Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. *Nature* 387:49-55

Allfrey VG, Faulkner R, Mirsky AE. (1964) Acetylation and methylation of histones and their possible roles in the regulation of RNA synthesis. *J Mol Biol* 196:397

Anderson, P. Mutagenesis. In *Caenorhabditis elegans: Modern Biological Analysis of an Organism*, San Diego: Academic Press 1995

Antequera F and Bird A. (1988) Unmethylated CpG islands associated with genes in higher plant DNA. *EMBO J* 7:2295-2299

Antequera F, Macleod D, Bird AP. (1989) Specific protection of methylated CpGs in mammalian nuclei. *Cell* 58:509-17

Antequera F, Tamame M, Villanueva JR, Santos T. (1984) DNA methylation in the fungi. *J Biol Chem* 259:8033-6

Arthur DC, Aaseng SM, Bloomfield CD. (1987) A new recurring chromosome abnormality in acute nonlymphocytic leukemia: del(2)(q23q33). *Cytogenet Cell Genet* 46:572

Ayer DE and Eisenman RN. (1993) A switch from Myc:Max to Mad:Max heterocomplexes accompanies monocyte/macrophage differentiation. *Genes Dev* 7:2110-9

Ayer DE, Lawrence QA, Eisenman RN. (1995) Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell* 80:767-76

Ball DJ, Gross DS, Garrard WT. (1983) 5-methylcytosine is localized in nucleosomes that contain histone H1. *Proc Natl Acad Sci U S A* 80:5490-4

Ballinger DG and Benzer S. (1989) Targeted gene mutations in *Drosophila*. *Proc Natl Acad Sci U S A* 86:9402-6

Bannister AJ, Kouzarides T. (1996) The CBP co-activator is a histone acetyltransferase. *Nature* 384:641-3

Barlev NA, Candau R, Wang L, Darpino P, Silverman N, Berger SL. (1995) Characterization of physical interactions of the putative transcriptional adaptor, ADA2, with acidic activation domains and TATA-binding protein. *J Biol Chem* 270:19337-44

Barlev NA, Poltoratsky V, Owen-Hughes T, Ying C, Liu L, Workman JL, Berger SL. (1998) Repression of GCN5 histone acetyltransferase activity via bromodomain-mediated binding and phosphorylation by the Ku-DNA-dependent protein kinase complex. *Mol Cell Biol* 18:1349-58

Bassett DE Jr, Boguski M, Hieter P. (1997) Identifying human homologs of cell cycle genes using dbEST and XREFdb. *Methods Enzymol* 283:128-40

- Batchelder C, Dunn MA, Choy B, Suh Y, Cassie C, Shim EY, Shin TH, Mello C, Seydoux G, Blackwell TK. (1999) Transcriptional repression by the *Caenorhabditis elegans* germ-line protein PIE-1. *Genes Dev* 13:202-12
- Bednarik DP, Cook JA, Pitha PM. (1990) Inactivation of the HIV LTR by DNA CpG methylation: evidence for a role in latency. *EMBO J* 9:1157-64
- Bednarik DP, Duckett C, Kim SU, Perez VL, Griffis K, Guenther PC, Folks TM. (1991) DNA CpG methylation inhibits binding of NF-kB proteins to the HIV-1 long terminal repeat cognate DNA motifs. *New Biol* 3:969-76
- Behe M and Felsenfeld G. (1981) Effects of methylation on a synthetic polynucleotide: the B--Z transition in poly(dG-m5dC).poly(dG-m5dC). *Proc Natl Acad Sci U S A* 78:1619-23
- Berger SL, Pina B, Silverman N, Marcus GA, Agapite J, Regier JL, Triezenberg SJ, Guarente L. (1992) Genetic isolation of ADA2: a potential transcriptional adaptor required for function of certain acidic activation domains. *Cell* 70:251-65
- Berkner KL and Folk WR. (1979) The effects of substituted pyrimidines in DNAs on cleavage by sequence-specific endonucleases. *J Biol Chem* 254:2551-60
- Bernards A, van Harten-Loosbroek N, Borst P. (1984) Modification of telomeric DNA in *Trypanosoma brucei*; a role in antigenic variation? *Nucleic Acids Res* 12:4153-70
- Bestor TH and Ingram VM. (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc Natl Acad Sci U S A* 80:5559-63
- Bestor TH. (1990) DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. *Philos Trans R Soc Lond B Biol Sci* 326:179-87
- Bhattacharya SK, Ramchandani S, Cervoni N, Szyf M. (1999) A mammalian protein with specific demethylase activity for mCpG DNA. *Nature* 397:579-83
- Bickmore WA, Oghene K, Little MH, Seawright A, van Heyningen V, Hastie ND. (1992) Modulation of DNA binding specificity by alternative splicing of the Wilms tumor wt1 gene transcript. *Science* 257:235-7
- Bird AP. (1986) CpG-rich islands and the function of DNA methylation. *Nature* 321:209-13
- Bird AP. (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* 8:1499-504
- Bird AP, Taggart MH, Smith BA. (1979) Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell* 17:889-901
- Blumenthal T and Thomas J. (1988) *Cis* and *trans* mRNA splicing in *C. elegans*. *Trends Genet* 4:305-8
- Boguski MS, Lowe TM, Tolstoshev CM. (1993) dbEST--database for "expressed sequence tags". *Nat Genet* 4:332-3

Borrow J, Stanton VP Jr, Andresen JM, Becher R, Behm FG, Chaganti RS, Civin CI, Distèche C, Dube I, Frischauf AM, Horsman D, Mitelman F, Volinia S, Watmore AE, Housman DE. (1996) The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. *Nat Genet* 14:33-41

Boyes J and Bird A. (1991) DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* 64:1123-34

Boyes J and Bird A. (1992) Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J* 11:327-33

Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, Temper V, Razin A, Cedar H. (1994) Sp1 elements protect a CpG island from de novo methylation. *Nature* 371:435-8

Braunstein M, Rose AB, Holmes SG, Allis CD, Broach JR. (1993) Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev* 7:592-604

Bresnick EH, John S, Hager GL. (1991) Histone hyperacetylation does not alter the positioning or stability of phased nucleosomes on the mouse mammary tumor virus long terminal repeat. *Biochemistry* 30:3490-7

Brenner S. (1974) The genetics of *Caenorhabditis elegans*. *Genetics* 77:71-94

Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, Allis CD. (1996) Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84:843-51

Bulyk ML, Gentalen E, Lockhart DJ, Church GM. (1999) Quantifying DNA-protein interactions by double-stranded DNA arrays. *Nat Biotechnol* 17:573-7

Buschhausen G, Graessmann M, Graessmann A. (1985) Inhibition of herpes simplex thymidine kinase gene expression by DNA methylation is an indirect effect. *Nucleic Acids Res* 13:5503-13

Buschhausen G, Wittig B, Graessmann M, Graessmann A. (1987) Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. *Proc Natl Acad Sci U S A* 84:1177-81

Busslinger M, Hurst J, Flavell RA. (1983) DNA methylation and the regulation of globin gene expression. *Cell* 34:197-206

Cairns BR, Lorch Y, Li Y, Zhang M, Lacomis L, Erdjument-Bromage H, Tempst P, Du J, Laurent B, Kornberg RD. (1996) RSC, an essential, abundant chromatin-remodeling complex. *Cell* 87:1249-60

Callebaut I, Courvalin JC, Mornon JP. (1999) The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. *FEBS Lett* 446:189-93

Campbell RD and Milner CM. (1993) MHC genes in autoimmunity. *Curr Opin Immunol* 5:887-93



- Campoy FJ, Meehan RR, McKay S, Nixon J, Bird A. (1995) Binding of histone H1 to DNA is indifferent to methylation at CpG sequences. *J Biol Chem* 270:26473-81
- Candau R, Zhou JX, Allis CD, Berger SL. (1997) Histone acetyltransferase activity and interaction with ADA2 are critical for GCN5 function in vivo. *EMBO J* 16:555-65
- The *C. elegans* Sequencing Consortium. (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282:2012-8
- Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. (1994) Green fluorescent protein as a marker for gene expression. *Science* 263:802-5
- Chaplin T, Ayton P, Bernard OA, Saha V, Della Valle V, Hillion J, Gregorini A, Lillington D, Berger R, Young BD. (1995) A novel class of zinc finger/leucine zipper genes identified from the molecular cloning of the t(10;11) translocation in acute leukemia. *Blood* 85:1435-41
- Chen ZJ and Pikaard CS. (1997) Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. *Genes Dev* 11:2124-36
- Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T, Weng S, Cherry JM, Botstein D. (1998) Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* 282:2022-8
- Chomczynski P and Sacchi N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156-9
- Chrivia JC, Kwok RP, Lamb N, Hagiwara M, Montminy MR, Goodman RH. (1993) Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature* 365:855-9
- Church GM and Gilbert W. (1984) Genomic sequencing. *Proc Natl Acad Sci U S A* 81:1991-5
- Cleary ML. (1992) Transcription factors in human leukaemias. *Cancer Surv* 15:89-104
- Collins M and Myers RM. (1987) Alterations in DNA helix stability due to base modifications can be evaluated using denaturing gradient gel electrophoresis. *J Mol Biol* 198:737-744
- Columbano-Green LM, Romain DR, Carter J, Crossen PE. (1990) t(2;14)(q23;q32.3) as the sole abnormality in a patient with acute nonlymphocytic leukemia (FAB-M4). *Cancer Genet Cytogenet* 48:255-7
- Cooper DN, Errington LH, Clayton RM. (1983) Variation in the DNA methylation pattern of expressed and nonexpressed genes in chicken. *DNA* 2:131-40
- Cooper DN and Youssoufian H. (1988) The CpG dinucleotide and human genetic disease. *Hum Genet* 78:151-5
- Cross SH, Meehan RR, Nan X, Bird A. (1997) A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nat Genet* 16:256-9

- Croston GE, Kerrigan LA, Lira LM, Marshak DR, Kadonaga JT. (1991) Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. *Science* 251:643-9
- Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY. (1995) Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci* 20:448-55
- Cusick ME, Lee KS, DePamphilis ML, Wassarman PM. (1983) Structure of chromatin at deoxyribonucleic acid replication forks: nuclease hypersensitivity results from both prenucleosomal deoxyribonucleic acid and an immature chromatin structure. *Biochemistry* 22:3873-84
- De Rubertis F, Kadosh D, Henchoz S, Pauli D, Reuter G, Struhl K, Spierer P. (1996) The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* 384:589-91
- Djabali M, Selleri L, Parry P, Bower M, Young BD, Evans GA. (1992) A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nat Genet* 2:113-8
- Dyck JA, Maul GG, Miller WH Jr, Chen JD, Kakizuka A, Evans RM. (1994) A novel macromolecular structure is a target of the promyelocyte-retinoic acid receptor oncoprotein. *Cell* 76:333-43
- Eden S, Hashimshony T, Keshet I, Cedar H, Thorne AW. (1998) DNA methylation models histone acetylation. *Nature* 394:842
- Epstein, HF and Shakes, DC. (1995) *Caenorhabditis elegans*: Modern Biological Analysis of an Organism, San Diego: Academic Press
- Feinberg AP and Vogelstein B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132:6-13
- Ferguson-Smith AC, Sasaki H, Cattanaach BM, Surani MA. (1993) Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* 362:751-5
- Fields C. (1990) Information content of *Caenorhabditis elegans* splice site sequences varies with intron length. *Nucleic Acids Res* 18:1509-12
- Finnegan EJ and Dennis ES. (1993) Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Res* 21:2383-8
- Finney M. (1987) Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806-11
- Fitch W. (1970) Distinguishing homologous from analogous proteins. *Syst. Zool.* 19:99-106
- Foss HM, Roberts CJ, Claeys KM, Selker EU. (1993) Abnormal chromosome behavior in *Neurospora* mutants defective in DNA methylation. *Science* 262:1737-41

Frank D, Keshet I, Shani M, Levine A, Razin A, Cedar H. (1991) Demethylation of CpG islands in embryonic cells. *Nature* 351:239-41

Gelmetti V, Zhang J, Fanelli M, Minucci S, Pelicci PG, Lazar MA. (1998) Aberrant recruitment of the nuclear receptor corepressor-histone deacetylase complex by the acute myeloid leukemia fusion partner ETO. *Mol Cell Biol* 18:7185-91

Georgakopoulos T, Thireos G. (1992) Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. *EMBO J* 11:4145-52

Gibbons RJ, Bachoo S, Picketts DJ, Aftimos S, Asenbauer B, Bergoffen J, Berry SA, Dahl N, Fryer A, Keppler K, Kurosawa K, Levin ML, Masuno M, Neri G, Pierpont ME, Slaney SF, Higgs DR. (1997) Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain. *Nat Genet* 17:146-8

Gossler A, Joyner AL, Rossant J, Skarnes WC. (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* 244:463-5

Goyon C and Faugeron G. (1989) Targeted transformation of *Ascobolus immersus* and de novo methylation of the resulting duplicated DNA sequences. *Mol Cell Biol* 9:2818-27

Goyon C, Nogueira TI, Faugeron G. (1994) Perpetuation of cytosine methylation in *Ascobolus immersus* implies a novel type of maintenance methylase. *J Mol Biol* 240:42-51

Grignani F, De Matteis S, Nervi C, Tomassoni L, Gelmetti V, Cioce M, Fanelli M, Ruthardt M, Ferrara FF, Zamir I, Seiser C, Grignani F, Lazar MA, Minucci S, Pelicci PG. (1998) Fusion proteins of the retinoic acid receptor- $\alpha$  recruit histone deacetylase in promyelocytic leukaemia. *Nature* 391:815-8

Gruenbaum Y, Cedar H, Razin A. (1982) Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature* 295:620-2

Gruenbaum Y, Stein R, Cedar H, Razin A. (1981a) Methylation of CpG sequences in eukaryotic DNA. *FEBS Lett* 124:67-71

Gruenbaum Y, Naveh-Man T, Cedar H, Razin A. (1981b) Sequence specificity of methylation in higher plant DNA. *Nature* 292:860-2

Gruenbaum Y, Szyf M, Cedar H, Razin A. (1983) Methylation of replicating and post-replicated mouse L-cell DNA. *Proc Natl Acad Sci U S A* 80:4919-21

Grunstein M. (1997) Histone acetylation in chromatin structure and transcription. *Nature* 389:349-52

Haynes SR, Dollard C, Winston F, Beck S, Trowsdale J, Dawid IB. (1992) The bromodomain: a conserved sequence found in human, *Drosophila* and yeast proteins. *Nucleic Acids Res* 20:2603

Hebbes TR, Clayton AL, Thorne AW, Crane-Robinson C. (1994) Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J* 13:1823-30

- Heinzel T, Lavinsky RM, Mullen TM, Soderstrom M, Laherty CD, Torchia J, Yang WM, Brard G, Ngo SD, Davie JR, Seto E, Eisenman RN, Rose DW, Glass CK, Rosenfeld MG. (1997) A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature* 387:43-8
- Hemesath TJ, Steingrimsson E, McGill G, Hansen MJ, Vaught J, Hodgkinson CA, Arnheiter H, Copeland NG, Jenkins NA, Fisher DE. (1994) microphthalmia, a critical factor in melanocyte development, defines a discrete transcription factor family. *Genes Dev* 8:2770-80
- Hendrich B, Bird A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* 18:6538-47
- Hendrich B, Hardeland U, Ng H-H, Jiricny J, Bird A. (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 401:301-04
- Hengartner MO, Horvitz HR. (1994) *C. elegans* cell survival gene ced-9 encodes a functional homolog of the mammalian proto-oncogene bcl-2. *Cell* 76:665-76
- Hodgkin J. (1994) Epigenetics and the maintenance of gene activity states in *Caenorhabditis elegans*. *Dev Genet* 15:471-7
- Hodgkin J, Horvitz HR, Brenner S. (1979) Nondisjunction mutants of the nematode *Caenorhabditis elegans*. *Genetics* 77:71-94
- Hollstein M, Shomer B, Greenblatt M, Soussi T, Hovig E, Montesano R, Harris CC. Somatic point mutations in the p53 gene of human tumors and cell lines: updated compilation. *Nucleic Acids Res* 1996 Jan 1;24(1):141-6
- Hope IA. (1994) PES-1 is expressed during early embryogenesis in *Caenorhabditis elegans* and has homology to the fork head family of transcription factors. *Development* 120:505-14
- Horlein AJ, Naar AM, Heinzel T, Torchia J, Gloss B, Kurokawa R, Ryan A, Kamei Y, Soderstrom M, Glass CK, et al. (1995) Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* 377:397-404
- Horz W and Altenburger W. (1981) Nucleotide sequence of mouse satellite DNA. *Nucleic Acids Res* 9:683-96
- Howlett SK and Reik W. (1991) Methylation levels of maternal and paternal genomes during preimplantation development. *Development* 113:119-27
- Hsieh CL. (1994) Dependence of transcriptional repression on CpG methylation density. *Mol Cell Biol* 14:5487-94
- Hsieh CL and Lieber MR. (1992) CpG methylated minichromosomes become inaccessible for V(D)J recombination after undergoing replication. *EMBO J* 11:315-25
- Huang LH, Farnet CM, Ehrlich KC, Ehrlich M. (1982) Digestion of highly modified bacteriophage DNA by restriction endonucleases. *Nucleic Acids Res* 10:1579-91

- Hug M, Silke J, Georgiev O, Rusconi S, Schaffner W, Matsuo K. (1996) Transcriptional repression by methylation: cooperativity between a CpG cluster in the promoter and remote CpG-rich regions. *FEBS Lett* 379:251-4
- Iguchi-Arigo SM and Schaffner W. (1989) CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev* 3:612-9
- Inouye S and Tsuji FI. (1994) Evidence for redox forms of the *Aequorea* green fluorescent protein. *FEBS Lett* 351:211-4
- Irniger S and Braus GH. (1994) Saturation mutagenesis of a polyadenylation signal reveals a hexanucleotide element essential for mRNA 3' end formation in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 91:257-61
- Ito T, Bulger M, Pazin MJ, Kobayashi R, Kadonaga JT. (1997) ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* 90:145-55
- Jansen G, Hazendonk E, Thijssen KL, Plasterk RH. (1997) Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. *Nat Genet* 17:119-21
- Jaenisch R. (1997) DNA methylation and imprinting: why bother? *Trends Genet* 13:323-9
- Jeanmougin F, Wurtz JM, Le Douarin B, Chambon P, Losson R. (1997) The bromodomain revisited. *Trends Biochem Sci* 22:151-3
- Jeddeloh JA, Stokes TL, Richards EJ. (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat Genet* 22:94-7
- Jiang J and Jacob HJ. (1998) EbEST: an automated tool using expressed sequence tags to delineate gene structure. *Genome Res* 8:268-75
- Johnson CA, Goddard JP, Adams RL. (1995) The effect of histone H1 and DNA methylation on transcription. *Biochem J* 305:791-8
- Johnson LM, Jayne PS, Kahn ES, Grunstein M. (1990) Genetic evidence for the interaction between SIR3 and histone H4 in the repression of the silent mating type loci in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 87:6286-6290
- Jones PA and Taylor SM. (1980) Cellular differentiation, cytidine analogs and DNA methylation. *Cell* 20:85-93
- Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 19:187-91
- Juttermann R, Li E, Jaenisch R. (1994) Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proc Natl Acad Sci U S A* 91:11797-801
- Kadosh D and Struhl K. (1997) Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell* 89:365-71

- Kafri T, Ariel M, Brandeis M, Shemer R, Urven L, McCarrey J, Cedar H, Razin A. (1992) Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev* 6:705-14
- Kamei Y, Xu L, Heinzl T, Torchia J, Kurokawa R, Gloss B, Lin SC, Heyman RA, Rose DW, Glass CK, Rosenfeld MG. (1996) A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell* 85:403-14
- Kass SU, Landsberger N, Wolffe AP. (1997) DNA methylation directs a time-dependent repression of transcription initiation. *Curr Biol* 7:157-65
- Kehle J, Beuchle D, Treuheit S, Christen B, Kennison JA, Bienz M, Muller J. (1998) dMi-2, a hunchback-interacting protein that functions in polycomb repression. *Science* 282:1897-900
- Keshet I, Lieman-Hurwitz J, Cedar H. (1986) DNA methylation affects the formation of active chromatin. *Cell* 44:535-43
- Kinzler KW and Vogelstein B. (1989) Whole genome PCR: application to the identification of sequences bound by gene regulatory proteins. *Nucleic Acids Res* 17:3645-53
- Kleff S, Andrulis ED, Anderson CW, Sternglanz R. (1995) Identification of a gene encoding a yeast histone H4 acetyltransferase. *J Biol Chem* 270:24674-7
- Koken MH, Saib A, de The H. (1995) A C4HC3 zinc finger motif. *C R Acad Sci III* 318:733-9
- Kolsto AB, Kollias G, Giguere V, Isobe KI, Prydz H, Grosveld F. (1986) The maintenance of methylation-free islands in transgenic mice. *Nucleic Acids Res* 14:9667-78
- Kozak M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44:283-92
- Kozak M. (1987) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J Mol Biol* 196:947-50
- Kozak M. (1996) Interpreting cDNA sequences: some insights from studies on translation. *Mamm Genome* 7:563-74
- Krause, M. (1995) Transcription and translation. In *Caenorhabditis elegans: Modern Biological Analysis of an Organism*, San Diego: Academic Press
- Krause M, Wild M, Rosenzweig B, Hirsh D. (1989) Wild-type and mutant actin genes in *Caenorhabditis elegans*. *J Mol Biol* 208:381-92
- Kruger W, Peterson CL, Sil A, Coburn C, Arents G, Moudrianakis EN, Herskowitz I. (1995) Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. *Genes Dev* 9:2770-9
- Laherty CD, Yang WM, Sun JM, Davie JR, Seto E, Eisenman RN. (1997) Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell* 89:349-56

- Lauster R, Trautner TA, Noyer-Weidner M. (1989) Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *J Mol Biol* 206:305-12
- Lee DY, Hayes JJ, Pruss D, Wolffe AP. (1993) A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* 72:73-84
- Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, Li E. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 122:3195-205
- Leonhardt H, Page AW, Weier HU, Bestor TH. (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 71:865-73
- Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* 69:905-14
- Levine A, Cantoni GL, Razin A. (1991) Inhibition of promoter activity by methylation: possible involvement of protein mediators. *Proc Natl Acad Sci U S A* 88:6515-8
- Levine A, Yeivin A, Ben-Asher E, Aloni Y, Razin A. (1993) Histone H1-mediated inhibition of transcription initiation of methylated templates in vitro. *J Biol Chem* 268:21754-9
- Li E, Beard C, Jaenisch R. (1993) Role for DNA methylation in genomic imprinting. *Nature* 366:362-5
- Li E, Bestor TH, Jaenisch R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69:915-26
- Lieb M and Rehmat S. (1997) 5-Methylcytosine is not a mutation hot spot in nondividing *Escherichia coli*. *Proc Natl Acad Sci U S A* 94:940-5
- Lindahl T and Nyberg B. (1974) Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* 13:3405-10
- Lock LF, Takagi N, Martin GR. (1987) Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell* 48:39-46
- Lu X, Meng X, Morris CA, Keating MT. (1998) A novel human gene, WSTF, is deleted in Williams syndrome. *Genomics* 54:241-9
- Lutterbach B, Westendorf JJ, Linggi B, Patten A, Moniwa M, Davie JR, Huynh KD, Bardwell VJ, Lavinsky RM, Rosenfeld MG, Glass C, Seto E, Hiebert SW. (1998) ETO, a target of t(8;21) in acute leukemia, interacts with the N-CoR and mSin3 corepressors. *Mol Cell Biol* 18:7176-84
- Macleod D, Ali RR, Bird A. (1998) An alternative promoter in the mouse major histocompatibility complex class III-A $\beta$  gene: implications for the origin of CpG islands. *Mol Cell Biol* 18:4433-43
- Macleod D, Charlton J, Mullins J, Bird AP. (1994) Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev* 8:2282-92

Makalowski W, Zhang J, Boguski MS. (1996) Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. *Genome Res* 6:846-57

Malagnac F, Wendel B, Goyon C, Faugeron G, Zickler D, Rossignol JL, Noyer-Weidner M, Vollmayr P, Trautner TA, Walter J. (1997) A gene essential for de novo methylation and development in *Ascomobolus* reveals a novel type of eukaryotic DNA methyltransferase structure. *Cell* 91:281-90

Manuelidis L. (1981) Consensus sequence of mouse satellite DNA indicates it is derived from tandem 116 basepair repeats. *FEBS Lett* 129:25-8

Matsuo K, Silke J, Georgiev O, Marti P, Giovannini N, Rungger D. (1998) An embryonic demethylation mechanism involving binding of transcription factors to replicating DNA. *EMBO J* 17:1446-53

McArthur M and Thomas JO. (1996) A preference of histone H1 for methylated DNA. *EMBO J* 15:1705-14

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58:499-507

Miller OJ, Schnedl W, Allen J, Erlanger BF. (1974) 5-Methylcytosine localised in mammalian constitutive heterochromatin. *Nature* 251:636-7

Miskimins WK, Roberts MP, McClelland A, Ruddle FH. (1985) Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferrin receptor gene. *Proc Natl Acad Sci U S A* 82:6741-4

Mizzen CA, Yang XJ, Kokubo T, Brownell JE, Bannister AJ, Owen-Hughes T, Workman J, Wang L, Berger SL, Kouzarides T, Nakatani Y, Allis CD. (1996) The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell* 87:1261-70

Monk M, Boubelik M, Lehnert S. (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* 99:371-82

Montgomery MK, Xu S, Fire A. (1998) RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 95:15502-7

Murray EJ and Grosveld F. J (1987) Site specific demethylation in the promoter of human gamma-globin gene does not alleviate methylation mediated suppression. *EMBO* 6:2329-35

Mushegian AR, Bassett DE Jr, Boguski MS, Bork P, Koonin EV. (1997) Positionally cloned human disease genes: patterns of evolutionary conservation and functional motifs. *Proc Natl Acad Sci U S A* 94:5831-6

Nagamine K, Peterson P, Scott HS, Kudoh J, Minoshima S, Heino M, Krohn KJ, Lalioti MD, Mullis PE, Antonarakis SE, Kawasaki K, Asakawa S, Ito F, Shimizu N. (1997) Positional cloning of the APECED gene. *Nat Genet* 17:393-8



- Nagy L, Kao HY, Chakravarti D, Lin RJ, Hassig CA, Ayer DE, Schreiber SL, Evans RM. (1997) Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. *Cell* 89:373-80
- Nan X, Campoy FJ, Bird A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* 88:471-81
- Nan X, Meehan RR, Bird A. (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res* 21:4886-92
- Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393:386-9
- Nan X, Tate P, Li E, Bird A. (1996) DNA methylation specifies chromosomal localization of MeCP2. *Mol Cell Biol* 16:414-21
- Navashin M. (1934) Chromosomal alterations caused by hybridization and their bearing upon certain general genetical problems. *Cytologia* 5:169-203
- Neddermann P and Jiricny J. (1993) The purification of a mismatch-specific thymine-DNA glycosylase from HeLa cells. *J Biol Chem* 268:21218-24
- Ng HH and Bird A. (1999) DNA methylation and chromatin modification. *Curr Opin Genet Dev* 9:158-63
- Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A. (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* 23:58-61
- Nichols J and Nimer SD. (1992) Transcription factors, translocations, and leukemia. *Blood* 80:2953-63
- Oliphant AR and Struhl K. (1987) The use of random-sequence oligonucleotides for determining consensus sequences. *Methods Enzymol* 155:568-82
- Ogawa E, Maruyama M, Kagoshima H, Inuzuka M, Lu J, Satake M, Shigesada K, Ito Y. (1993) PEBP2/PEA2 represents a family of transcription factors homologous to the products of the *Drosophila* runt gene and the human AML1 gene. *Proc Natl Acad Sci U S A* 90:6859-63
- Ogryzko VV, Schiltz RL, Russanova V, Howard BH, Nakatani Y. (1996) The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* 87:953-9
- Okano M, Xie S, Li E. (1998a) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* 19:219-20
- Okano M, Xie S, Li E. (1998b) Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res* 26:2536-40
- Panning B and Jaenisch R. (1996) DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes Dev* 10:1991-2002
- Paro R. (1993) Mechanisms of heritable gene repression during development of *Drosophila*. *Curr Opin Cell Biol* 5:999-1005

- Paro R. (1990) Imprinting a determined state into the chromatin of *Drosophila*. Trends Genet 6:416-21
- Pawlak A, Bryans M, Jost JP. (1991) An avian 40 KDa nucleoprotein binds preferentially to a promoter sequence containing one single pair of methylated CpG. Nucleic Acids Res 19:1029-34
- Pays E, Delauw MF, Laurent M, Steinert M. (1984) Possible DNA modification in GC dinucleotides of *Trypanosoma brucei* telomeric sequences; relationship with antigen gene transcription. Nucleic Acids Res 12:5235-47
- Pazin MJ, Bhargava P, Geiduschek EP, Kadonaga JT. (1997) Nucleosome mobility and the maintenance of nucleosome positioning. Science 276:809-12
- Pazin MJ and Kadonaga JT. (1997) What's up and down with histone deacetylation and transcription? Cell 89:325-8
- Pengue G and Lania L. (1996) Kruppel-associated box-mediated repression of RNA polymerase II promoters is influenced by the arrangement of basal promoter elements. Proc Natl Acad Sci U S A 93:1015-20
- Peoples RJ, Cisco MJ, Kaplan P, Francke U. (1998) Identification of the WBSCR9 gene, encoding a novel transcriptional regulator, in the Williams-Beuren syndrome deletion at 7q11.23. Cytogenet Cell Genet 82:238-46
- Peterson CL, Dingwall A, Scott MP. (1994) Five SWI/SNF gene products are components of a large multisubunit complex required for transcriptional enhancement. Proc Natl Acad Sci U S A 91:2905-8
- Petrij F, Giles RH, Dauwerse HG, Saris JJ, Hennekam RC, Masuno M, Tommerup N, van Ommen GJ, Goodman RH, Peters DJ, et al . (1995) Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. Nature 376:348-51
- Plasterk RHA. (1995) Reverse genetics: From gene sequence to mutant worm. In *Caenorhabditis elegans: Modern Biological Analysis of an Organism*, San Diego: Academic Press
- Pollack Y, Stein R, Razin A, Cedar H. (1980) Methylation of foreign DNA sequences in eukaryotic cells. Proc Natl Acad Sci U S A 77:6463-7
- Posfai J, Bhagwat AS, Posfai G, Roberts RJ. (1989) Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res 17:2421-35
- Pradhan S and Adams RL. (1995) Distinct CG and CNG DNA methyltransferases in *Pisum sativum*. Plant J 7:471-81
- Prasad R, Leshkowitz D, Gu Y, Alder H, Nakamura T, Saito H, Huebner K, Berger R, Croce CM, Canaani E. (1994) Leucine-zipper dimerization motif encoded by the AF17 gene fused to ALL-1 (MLL) in acute leukemia. Proc Natl Acad Sci U S A 91:8107-11
- Proffitt JH, Davie JR, Swinton D, Hattman S. (1984) 5-Methylcytosine is not detectable in *Saccharomyces cerevisiae* DNA. Mol Cell Biol 4:985-8

- Quaderi NA, Meehan RR, Tate PH, Cross SH, Bird AP, Chatterjee A, Herman GE, Brown SD. (1994) Genetic and physical mapping of a gene encoding a methyl CpG binding protein, Mecp2, to the mouse X chromosome. *Genomics* 22:648-51
- Rabbitts TH. (1994) Chromosomal translocations in human cancer. *Nature* 372:143-9
- Rae PM, Steele RE. (1979) Absence of cytosine methylation at C-C-G-G and G-C-G-C sites in the rDNA coding regions and intervening sequences of *Drosophila* and the rDNA of other insects. *Nucleic Acids Res* 6:2987-95
- Rinderle C, Christensen HM, Schweiger S, Lehrach H, Yaspo ML. (1999) AIRE encodes a nuclear protein co-localizing with cytoskeletal filaments: altered sub-cellular distribution of mutants lacking the PHD zinc fingers. *Hum Mol Genet* 8:277-90
- Riggs AD and Pfeifer GP. (1992) X-chromosome inactivation and cell memory. *Trends Genet* 8:169-74
- Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL. (1996) Demethylation-induced developmental pleiotropy in *Arabidopsis*. *Science* 273:654-7
- Rosenzweig B, Liao LW, Hirsh D. (1983) Target sequences for the *C. elegans* transposable element Tc1. *Nucleic Acids Res* 11:7137-40
- Rothnie HM, McCurrach KJ, Glover LA, Hardman N. (1991) Retrotransposon-like nature of Tp1 elements: implications for the organisation of highly repetitive, hypermethylated DNA in the genome of *Physarum polycephalum*. *Nucleic Acids Res* 19:279-86
- Rundlett SE, Carmen AA, Kobayashi R, Bavykin S, Turner BM, Grunstein M. (1996) HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proc Natl Acad Sci U S A* 93:14503-8
- Ruppert S, Wang EH, Tjian R. (1993) Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation. *Nature* 362:175-9
- Saha V, Chaplin T, Gregorini A, Ayton P, Young BD. (1995) The leukemia-associated-protein (LAP) domain, a cysteine-rich motif, is present in a wide range of proteins, including MLL, AF10, and MLLT6 proteins. *Proc Natl Acad Sci U S A* 92:9737-41
- Sanchez-Garcia I and Rabbitts TH. (1994) The LIM domain: a new structural motif found in zinc-finger-like proteins. *Trends Genet* 10:315-20
- Selker EU. (1998) Trichostatin A causes selective loss of DNA methylation in *Neurospora*. *Proc Natl Acad Sci U S A* 95:9430-5
- Sessarego M, Grammenu S, Scarra GB, Ajmar F. (1979) A case of chronic myelogenous leukemia with unusual chromosomal abnormality. *Leuk Res* 3:271-5
- Scharf JM, Endrizzi MG, Wetter A, Huang S, Thompson TG, Zerres K, Dietrich WF, Wirth B, Kunkel LM. (1998) Identification of a candidate modifying gene for spinal muscular atrophy by comparative genomics. *Nat Genet* 20:83-6
- Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW. (1998) Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol* 16:301-6

Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tome P, Aggarwal A, Bajorek E, Bentolila S, Birren BB, Butler A, Castle AB, Chiannilkulchai N, Chu A, Clee C, Cowles S, Day PJ, Dibling T, Drouot N, Dunham I, Duprat S, East C, Hudson TJ, et al. (1996) A gene map of the human genome. *Science* 274:540-6

Shemer R, Kafri T, O'Connell A, Eisenberg S, Breslow JL, Razin A. (1991) Methylation changes in the apolipoprotein AI gene during embryonic development of the mouse. *Proc Natl Acad Sci U S A* 88:11300-4

Sheridan PL, Mayall TP, Verdin E, Jones KA. (1997) Histone acetyltransferases regulate HIV-1 enhancer activity in vitro. *Genes Dev* 11:3327-40

Shi Y, Seto E, Chang LS, Shenk T. (1991) Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* 67:377-88

Shi Y and Mello C. (1998) A CBP/p300 homolog specifies multiple differentiation pathways in *Caenorhabditis elegans*. *Genes Dev* 12:943-55

Shrivastava A and Calame K. (1994) An analysis of genes regulated by the multi-functional transcriptional regulator Yin Yang-1. *Nucleic Acids Res* 22:5151-5

Silverman N, Agapite J, Guarente L. (1994) Yeast ADA2 protein binds to the VP16 protein activation domain and activates transcription. *Proc Natl Acad Sci U S A* 91:11665-8

Simmen MW, Leitgeb S, Charlton J, Jones SJ, Harris BR, Clark VH, Bird A. (1999) Nonmethylated transposable elements and methylated genes in a chordate genome. *Science* 283:1164-7

Simpson VJ, Johnson TE, Hammen RF. (1986) *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res* 14:6711-9

Sobel RE, Cook RG, Perry CA, Annunziato AT, Allis CD. (1995) Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. *Proc Natl Acad Sci U S A* 92:1237-41

Solage A and Cedar H. (1978) Organization of 5-methylcytosine in chromosomal DNA. *Biochemistry* 17:2934-8

Solari F, Bateman A, Ahringer J. (1999) The *Caenorhabditis elegans* genes *egl-27* and *egr-1* are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. *Development* 126:2483-94

Southern EM. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503-17

Spieth J, Brooke G, Kuersten S, Lea K, Blumenthal T. (1993) Operons in *C. elegans*: polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell* 73:521-32

Stein R, Razin A, Cedar H. (1982) In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. *Proc Natl Acad Sci U S A* 79:3418-22

- Strauss F and Varshavsky A. (1984) A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. *Cell* 37:889-901
- Sulston JE, Schierenberg E, White JG, Thomson JN. (1983) The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev Biol* 100:64-119
- Sun XH and Baltimore D. (1991) An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. *Cell* 64:459-70
- Szyf M, Milstone DS, Schimmer BP, Parker KL, Seidman JG. (1990) *cis* modification of the steroid 21-hydroxylase gene prevents its expression in the Y1 mouse adrenocortical tumor cell line. *Mol Endocrinol* 4:1144-52
- Tamkun JW, Deuring R, Scott MP, Kissinger M, Pattatucci AM, Kaufman TC, Kennison JA. (1992) *brahma*: a regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* 68:561-72
- Tanaka T, Tanaka K, Ogawa S, Kurokawa M, Mitani K, Nishida J, Shibata Y, Yazaki Y, Hirai H. (1995) An acute myeloid leukemia gene, AML1, regulates hemopoietic myeloid cell differentiation and transcriptional activation antagonistically by two alternative spliced forms. *EMBO J* 14:341-50
- Tate P, Skarnes W, Bird A. (1996) The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. *Nat Genet* 12:205-8
- Taunton J, Hassig CA, Schreiber SL. (1996) A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* 272:408-11
- Tazi J and Bird A. (1990) Alternative chromatin structure at CpG islands. *Cell* 60:909-20
- Thoma F, Koller T, Klug A. (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol* 83:403-27
- Travers A. (1999) An engine for nucleosome remodeling. *Cell* 96:311-4
- Tsukiyama T, Becker PB, Wu C. (1994) ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* 367:525-32
- Tsukiyama T, Daniel C, Tamkun J, Wu C. (1995) ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* 83:1021-6
- Tweedie S, Charlton J, Clark V, Bird A. (1997) Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol* 17:1469-75
- Vardimon L, Kressmann A, Cedar H, Maechler M, Doerfler W. (1982) Expression of a cloned adenovirus gene is inhibited by in vitro methylation. *Proc Natl Acad Sci U S A* 79:1073-7

- Varga-Weisz PD, Wilm M, Bonte E, Dumas K, Mann M, Becker PB. (1997) Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* 388:598-602
- Vidal M and Gaber RF. (1991) RPD3 encodes a second factor required to achieve maximum positive and negative transcriptional states in *Saccharomyces cerevisiae*. *Mol Cell Biol* 11:6317-27
- Vongs A, Kakutani T, Martienssen RA, Richards EJ. (1993) *Arabidopsis thaliana* DNA methylation mutants. *Science* 260:1926-8
- Wade PA, Geggion A, Jones PL, Ballestar E, Aubry F, Wolffe AP. (1999) Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet* 23:62-66
- Wade PA, Jones PL, Vermaak D, Wolffe AP. (1998) A multiple subunit Mi-2 histone deacetylase from *Xenopus laevis* cofractionates with an associated Snf2 superfamily ATPase. *Curr Biol* 8:843-6
- Wakefield RID, Smith BO, Nan X, Free A, Soteriou A, Uhrin D, Bird AP, Barlow P. (1999) The solution structure of the domain from MeCP2 that binds to methylated DNA. *J Mol Biol* 291:1055-1065
- Walsh CP and Bestor TH. (1999) Cytosine methylation and mammalian development. *Genes Dev* 13:26-34
- Wang RY, Zhang XY, Khan R, Zhou YW, Huang LH, Ehrlich M. (1986) Methylated DNA-binding protein from human placenta recognizes specific methylated sites on several prokaryotic DNAs. *Nucleic Acids Res* 14:9843-60
- Watt F and Molloy PL. (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev* 2:1136-43
- Weintraub H. (1984) Histone-H1-dependent chromatin superstructures and the suppression of gene activity. *Cell* 38:17-27
- White JG, Southgate E, Thomson JN, Brenner S. (1987) The structure of the nervous system of *Caenorhabditis elegans*. *Philos Trans R Soc Lond* 314:1-340
- Wilkinson CR, Bartlett R, Nurse P, Bird AP. (1995) The fission yeast gene *pmt1+* encodes a DNA methyltransferase homologue. *Nucleic Acids Res* 23:203-10
- Wilson R, Ainscough R, Anderson K, Baynes C, Berks M, Bonfield J, Burton J, Connell M, Copsey T, Cooper J, et al. (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* 368:32-8
- Wolffe AP. (1989) Dominant and specific repression of *Xenopus* oocyte 5S RNA genes and satellite I DNA by histone H1. *EMBO J* 8:527-37
- Wolffe AP, Wong J, Pruss D. (1997) Activators and repressors: making use of chromatin to regulate transcription. *Genes Cells* 2:291-302
- Wood WB. (1988) The Nematode *Caenorhabditis elegans*. Cold spring Harbor Laboratory Press.

- Xu W, Edmondson DG, Roth SY. (1998) Mammalian GCN5 and P/CAF acetyltransferases have homologous amino-terminal domains important for recognition of nucleosomal substrates. *Mol Cell Biol* 18:5659-69
- Yandell MD, Edgar LG, Wood WB. (1994) Trimethylpsoralen induces small deletion mutations in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 91:1381-5
- Yang WM, Inouye C, Zeng Y, Bearss D, Seto E. (1996) Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. *Proc Natl Acad Sci U S A* 93:12845-50
- Yang XJ, Ogryzko VV, Nishikawa J, Howard BH, Nakatani Y. (1996) A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* 382:319-24
- Yano T, Nakamura T, Blechman J, Sorio C, Dang CV, Geiger B, Canaani E. (1997) Nuclear punctate distribution of ALL-1 is conferred by distinct elements at the N terminus of the protein. *Proc Natl Acad Sci U S A* 94:7286-91
- Ye F and Singer ER. (1996) RIGS (repeat-induced gene silencing) in *Arabidopsis* is transcriptional and alters chromatin configuration. *Proc Natl Acad Sci U S A* 93:10881-10886
- Yisraeli J, Frank D, Razin A, Cedar H. (1988) Effect of in vitro DNA methylation on beta-globin gene expression. *Proc Natl Acad Sci U S A* 85:4638-42
- Yoder JA, Walsh CP, Bestor TH. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13:335-40
- Yoshida M, Kijima M, Akita M, Beppu T. (1990) Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *J Biol Chem* 265:17174-9
- Zambrowicz BP, Friedrich GA, Buxton EC, Lilleberg SL, Person C, Sands AT. (1998) Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* 392:608-11
- Zhang Y, Iratni R, Erdjument-Bromage H, Tempst P, Reinberg D. (1997) Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell* 89:357-64
- Zhang Y, LeRoy G, Seelig HP, Lane WS, Reinberg D. (1998) The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* 95:279-89
- Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* 13:1924-35
- Zwaal RR, Broeks A, van Meurs J, Groenen JT, Plasterk RH. (1993) Target-selected gene inactivation in *Caenorhabditis elegans* by using a frozen transposon insertion mutant bank. *Proc Natl Acad Sci U S A* 90:7431-5